

QY 904 INKPEETAKKDRATNSKATTPKOKPTKAPKKSTKKKTMRYRKK----- 954
 Db 1194 -LNKKAPAKETEK-----PKKEVSKKPPKPTPEPKP-AAKKKKPPMEDDDDEPE 1243
 QY 955 ---TTTPPKKSTMP 967
 Db 1244 ADFTMPAKKPDTEP 1259

RESULT 11
 09LIE8
 ID 09LIE8 PRELIMINARY; PRT: 1480 AA.
 AC 09LIE8;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE SMILAIRY TO CELL WALL-PLASMA MEMBRANE LINKER PROTEIN.
 OS Arabidopsis thaliana (mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC euroids II; Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_Taxid=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-COLUMBIA;
 RA Kaneo T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-COLUMBIA;
 RX PubMed-10907853;
 RA Nakamura Y.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
 RT TAC and BAC clones."
 RL DNA Res. 7:217-221(2000).
 DR EMBL: AP001306; BAB03062.1;
 DR Interpro: IPR002965; P. rich-extendn.
 DR PRINTS: PR01217; PRICHTXTNSN.
 SQ SEQUENCE 1480 AA; 147153 MW; D1ACOC79F155E732 CRC64;

Query Match 11.38; Score 800.5; DB 10; Length 1480;
 Best Local Similarity 29.68; Pred. No. 1.2e-45;
 Matches 307; Conservative 54; Mismatches 505; Indels 171; Gaps 45;

QY 116 KKKPKRPVVDGSGDNDGKVTTPDSTQHNKVS-----PKTTAKPINRPSL 171
 Db 117 KRPKPKPPNKK-----PKNPKRPKPKPKNPKRPKPNPKPKPPKPPSV 167
 QY 172 PPNDSKETSIVNKETTVETKTTNKQSTGDKKETSKEKQISIKTSKDLAPT 231
 Db 168 KPPSTPK-----PRTTNPPSTPQPPPKPPPCPTTPPVASBPVATP 211
 QY 232 SKV--LAKPTKATTTGPAITPKKEPT--TTP--KEBASTPKKEPTTTIKSATP 285
 Db 212 TQMPPIATP-PIAKSPVATPPIATPPIATPPIATPPIATPPIATPPIATP 270
 QY 286 KEPAPTT-----KSAPTKEKAPPTT-----KEBAPTTPKEAPPTTKE 326
 Db 271 PVAPPTTNPPISKPPVTP-----PTTTPPIAKPPIATPPISTPPATPPAATPPTTTP 326
 QY 327 PA-----PTTKSATPTP--KEBAPTTPKKAP--TTKEBAPTTPKEPTTPPTKEAP 376
 Db 327 PAKPVAISIVT--PVPVIAQPPVATPPTATPVPVATPPIATPPIATPPIATPPIATP 384
 QY 377 TTKEPATTTP--KEBAPTTPKKAPPTTPKEBAPTTPKEBAPTTP-----KEBPTTPKEAP 431
 Db 385 VAVPPTATSPKTPPKAPKPVATPPIAKSPVATPPIATPPIATPPIATPPIATPPTTTP 444
 QY 432 TTTSKAPTTPKEAPTTTTSAPTTPKEP--SPTTKEBAPTTPKEAP--TTTPKKAPT 486
 Db 445 TAT--PVAKPPVETPPIATPPTAPKPISTPPISKPVATPPIATPPIATPPTTTPVAKPVA 501

QY 487 TPKEBAPTTPKEP--APTTPKKAPATP--KEBAPTTPKEBAPT-----PKLTP--T 534
 Db 502 TTPPLAIPVAKPPVTPPTATPPIATPPIATPPIAKSPVATPPTATPPIATPPIATPPTT 561
 QY 535 TPKEBAPTTPKEBAPTTPKEBAPTTP--PEEPTTPPEEAPPTPKAAPT--PKEBAPT 590
 Db 562 TPTATPPVAKPPVATPPIATPPIATPPIATPPIATPPIATPPIATPPIATPPIATPPIATP 621
 QY 591 TPKEBAPTTPKEBAPTTPKEBAPT--TTPKGAAPT--LKEBAPTTPKKAPKELAPPTT 645
 Db 622 TPIATPPIAKPPVATPPTATPPIATPPIATSPVAKPPVATPPIATPPIATPPIATPPIATP 661
 QY 646 KEPTSTSDKAPPTPKGTAP--TPKEBAPTTPKEBAPTTPKGAAPTTPKEBAPTTPKPK 704
 Db 682 KPPVAT--PTATPPIATPPIATPPIATPPIATPPIATPPIATPPIATPPIATPPIATPPIAT 736
 QY 705 APKELAPPTTGPST--TSDKAPPTPKETAAPT--TPKEBAPTTPKKAPATP-----E 755
 Db 737 MPPIATPPIAKPPIATPPIAIPVAKPPVTPPTATPPIATPPIATPPIATPPIATPPIATP 796
 QY 756 TPPTSEVSPPTTPKPTTH--KSPDESTPELSABPTPKALENSPKKEGVPPTTPPA 813
 Db 797 TPIAKPPVATPPTATPPIATPPIAKPPVATPPIATPPIATPPIATPPIATPPIATPPIATP 856
 QY 814 TPKEBPT--TKAKDTERDLTPTETTAAPKMTKETATTEKTESKITATTTQVSTTT 872
 Db 857 TPTTTPPAKPPVATPPIATP--IAKPPVATPPIATPPIATPPIATPPIATPPIATPPIATP 914
 QY 873 QDTPPEITTLKTTTLAPKTTTKT-----ITTEIMNKEETAKKDRATNSKATTPK 928
 Db 915 PAKPVAIPPIATPPIAKPPVATPPIATPPIATPPIATPPIATPPIATPPIATPPIATPPIATP 971
 QY 929 QKPTKAPKPTSTKPKTPKRVKPKPTTPPKMTSTPELMP-----TSKTAE--- 977
 Db 972 KPEPTTP--PTAP--PVAAMPPIATP--PTAKPVAIPPIATPPIATPPIATPPIATPPIATP 1025
 QY 978 -----MLQTTTPNOTPVSKL--VEVNPKSDAGAGETGPHMLRPHVFEVETPMD 1030
 Db 1026 VLPPIAKPPVETSPVATPPIATPPIATPPIATPPIATPPIATPPIATPPIATPPIATPPIATP 1070
 QY 1031 YLPRVNOGIIINPMLS 1047
 Db 1071 ATPPVATNPPTAMPPIAT 1087

RESULT 12
 041805
 ID 041805 PRELIMINARY; PRT: 1188 AA.
 AC 041805;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE EXTENSIN-LIKE PROTEIN PRECURSOR.
 OS Zea mays (Maize).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
 OC Panicoidae; Andropogoneae; Zea.
 OC NCBI_Taxid=4577;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-B73; TISSUE-POLLEN;
 RA Rudinstein A.L., Broadwater A.H., Lowrey K., Bedinger P.A.;
 RL Submitted (JUN-1994) to the EMBL/Genbank/DBJ databases.
 DR EMBL: Z34465; CA84230.1;
 DR Mendel: 14346; Zeama; 2368; 14346.
 DR Interpro: IPR001611; LRR.
 DR Interpro: IPR003592; LRR_out.
 DR Pfam: PF00560; LRR; 3.
 DR SMART: SM00370; LRR; 4.
 DR signal.
 FT SIGNAL.
 SQ SEQUENCE 1 27 POTENTIAL.
 1188 AA; 120980 MW; 2C77C7F8D7130149 CRC64;

```

QY 664 TATPTP-----KEPAATTKKEAPATTPK--GTAPTTLKKEAPATTPKPKAPKELAPTT 713
DB 994 LPPAPVSSPPPVKSSPPPTPVSSPPPTPKLPPPTPVSSPPPTKXKLPP--APVS 1050
QY 714 TKGPTSTSDKPAPTT--PKETAPTPK-----EPATTPKPAATTPETPPPTTSEVSTP 767
DB 1051 SPPPVKSSPPAPVSLPPPTPKRSPPTRVSSPRVVKCCPPTLVSSPPAPKSLPPP 1110
QY 768 TTTKEPTTIHKSDESTPELSAETPKALENSKEGVPTTKTPATTKEMTTAAKDKT 827
DB 1111 TPVSSP-----PPEVKSSPPPTPVSSPPAPKSSPPPTVSSPP-- 1150
QY 828 ERDLKTP-----ETTAAPKMKETATTEKTESKITATTAQVSTTQDTTPPKITT 862
DB 1151 --ELKSSPPAPVSSPPAPKSSPPAPVSLPPEVKSSPPAPISPPAPKSP----- 1203
QY 883 LKTTLPKVTYTTTKKTTTTEIMNKPBEETAKPKDRATNSKATTPKPKPKAKKPTSTK 942
DB 1204 -----PPPA-----PMSSLPPPVKSSPPAPVSSPPPTPKSPPPAPIS 1242
QY 943 KPKTMPRVRKKTTPTPPKMTSTPELNPTRISLEAMLOTTTRPNQTPNSKLEVNPKSE 1002
DB 1243 SPPAP--VKPPSLPPAPVSSPPAPV-----TSAP-----PKKE 1275
QY 1003 DAGAEGETPHMLRPVHVEVTPDMDYLPRVNOGIIINPLMSDE 1049
DB 1276 EDSTP-----PPAPLPPPSFNDIILPPIMANK 1303

RESULT 10
Q20007 ID 020007 PRELIMINARY; PRT; 1274 AA.
AC 020007:
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMblrel. 08, Last annotation update)
DE COSMID F3545.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2; PubMed=7906398;
RX MEDLINE=94150718;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M., Coulson A.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fulton L.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Ritken L., Roopa A., Saunders D., Showkhen R.,
RA Smailson N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sprat J., Wohlman P.;
RA *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.
RT Nature 368:32-38(1994).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Leimbach D.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
[4]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;

```

RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U46675; AAB52641.1; C;
 SQ SEQUENCE 1274 AA; 138065 MW; 07F6B0D292C1799F CMC64;

```

Query Match 11.4%; Score 802; DB 5; Length 1274;
Best local similarity 28.0%; Pred. No. 8,5e-66;
Matches 290; Conservative 102; Mismatches 414; Indels 230; Gaps 56;

QY 112 KNRTRK-KP-----KPPVDEAGSGLDNGD-----FKTTPDSTQTNKYSTS 156
DB 274 KNPTRKMPWEDETPEVEKPPVKKAPVLLKKDAPAKARDSPSKAAKVPK 333
QY 157 KRTTAKPI-NRPSLPPNSDTSKETSLSVKNKETTVEKETTNNKQTSIDGKKTSAK 215
DB 334 SPVPPPTPVKNVKKYKPPWVEVDEPABEVKPSAKPKTPVLLKKEPEPSSTPSDPS 393
QY 216 ETOSIEKTSKADLAPTSKVLAKPTPKAE-----TTTKGA-----LUTP 254
DB 394 PKKAAPAVPRDSSPKKATPLQADEKAEVPTPVKNVKKYKPPWVEVDEDPVEEVKOP 453
QY 255 KEPTPTP-----KEPASTPKEP-----TP-TIKSAPTPKEPAPTTTSAVTPKREPAP 305
DB 454 EAPAKTPVLRKEPAKADPAKPAISKTPETPEKDPVPRDSSPKVAAKPDASAQADA- 512
QY 306 TTTKEPA-----PTTPKEPAPTTTKEPAP-----TTKSAPTTP 339
DB 513 TPVKNVKKWRPWEDETPADDSKPTDAKKTPLSLAKKDPAPAKESLKPADTKAPAKP 572
QY 340 KEP-----APTTKRPAPTTPK-----PAPTPKEPTTPTPKEPAPTTTKEPAPTTK-EP 389
DB 573 RDPSPKVAAPTADEKTPVLAKEKAPGADSKTEPEKSKPRDPSKKAVPAKVPKTEV 632
QY 390 APTAPKKAP-----TTKEPAPTTTKEPAPTTTKEP-----SPTPKRPAPTTTKSAPT 439
DB 633 APAVKKPELISKPKDAPAKKAPNSPVVP-PTPVKNVKKWRPWEDEDPAPKVSLSDE 691
QY 440 TTKEPAPTTTKSAPTTT-KEPSPTTTKEPAPTTPK--EPAPTTPKKAP--TTTKEPAPT 494
DB 692 PEKK-TPVLLAKKAPTCKDSEADDPVSGPSSKPKLAKKAPRDRPSWKAAVPIPAK 750
QY 495 TPKPAPTTTTPKAPTA--PEKAPTTTKEPAPTTTTPKLTPTP-----EKLAPTTEKPA 548
DB 751 T--EVPRAVYKPKPEPVAKSDSPSKAK--AEPNSP--VVPPTPVKNVKKWRPWEDDA 805
QY 549 PTPEELAPTTPEP-----TPTPEEPAPTTTAKAAPTTPKEPAPTTTKEPAPTTT-PE 602
DB 806 PAEPPVNPPEPKKPPVLAKEKTPVAKPRDPSPKKAVPAKSTKTAPRVSVKPPVSKPE 865
QY 603 PAPTPKETAATTPKGTAPTTLKEPAPTTTTPKPAKELAPT--TTKEPTSTTSOKPAPTTP 661
DB 866 PSPKKAEPNSVVP-----PTPVKNVKKW--KPWEDEDEPTTEVKKPSE--PEKTPVLA 918
QY 662 KGTAPTTKEPAPTTTKEPAPTTTTPKGTAPTTLKEPAPT-----TPKKPAP-----KEL 709
DB 919 K-KEPEKPD-APKVAAPRDPSPKAVPE--KPEKAVAAKPRLSKKAIPIPANQEA 974
QY 710 APTTTKSPTS-----TTSOKP-----APTTPKET-----APTTPEPAPTTTTPKAPT 752
DB 975 PPTPVKNVKKWRPWEDEDEPABEVSAPEKKTTPVLAKEKAPKAPRDP--SPKKAPV 1031
QY 753 PPTPEPTTSEVSTPTTKEPTTIHKS-----DESTPELSA-EP-----TPKALENSPKEPG 804
DB 1032 AAK-FDPKIPKV-PTPVKNVKKWRPWEDEDESEVSAPEKKTTPVLAKEKAPKAPA 1089
QY 805 V-----PTTKTPAPT-----KPEMTTAKDKTTERDLRTTPETT--TAAPK 843
DB 1090 TKPDSEAADPVSGPTSKPKLSKAPVHKPKPTTDPKDKLKPSPAKPKAPAPAPAPK 1149
QY 844 MKKETATTEKTESKITATTAQVSTTQDTTPPKITTLTKTTLTAPVTTTKKTTTTE 903
DB 1150 KKKPVMDDDPDEADFTVPAPSKKPDTEDPADPLG-----GPKTKDPK----- 1193

```

QY 227 DLAPSTVLAKPTPKAETTTGSPALTTKPEPTTPKEPASTPKPEPTTTIKSAPTTK 286
 DB 478 TVAPTKETTVAPT---EETTVASTETTVYAPTEETTVAPAEETTPYEET---TVAPTEET 533
 QY 287 EPAPT--TTKAPPTPKAPAP--TTKEPAPTTKAPAPTTTKEPAPTTKSAPT----- 337
 DB 534 TVAPTEETTVAPTEETTVAPTEETTVAPAEETTPYEETTVAPTEETTVAPTEETTVAP 592
 QY 338 ---TKEPAPTTPKKAPPTPKAPPTPKAPPTPKAPPTPKAPPTPKAPPTPKAPPTPK 376
 DB 593 PTEETTVAPTEETTVAPAEETTPYEETTVAPTEETTVAPTEETTVAPTEETTVAPTEET 652
 QY 377 TTKEPAPTTPKPE-----APTAKKAPAPTTKAPAPTTKAPAPTTKAPAPTTKAP 428
 DB 653 TVAPAEETTPYEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAP 712
 QY 429 P-----APT--TKSAPT-----TKEP-----APT--TKSAPT--TKEPSPPT 462
 DB 713 PTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTE 772
 QY 463 TTKEP-----APTTPKEPAPTTPKKAPAPTTKAPAPTTKAPAPTTKAPAPTTKAP 514
 DB 773 TVPEPTGETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEET 832
 QY 515 PAPTPKETAAPTTPKKAPPTPKAPPTPKAPPTPKAPPTPKAPPTPKAPPTPKAPPTPK 571
 DB 833 PTEET---TTPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAP 890
 QY 572 ---APT-----TPKAAAPNTPKAPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAPTT 610
 DB 891 KETTVAPAEETTVASTETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEET 950
 QY 611 TAPTPKGTAP--TLKAPAPTTKAPAPK-ELAPT--TTKEPTSTSDKAPAPTTKAP 665
 DB 951 YAPTEETTVAPTEETTVAPAEETTPYEETTVAPTEETTVAPTEETTVAPTEETTVAPTE 1010
 QY 666 PTPKAPAPTTPKPE-----APTTPKGTAP-----APTTPKAPAPTTPKPE 701
 DB 1011 EETTVAPAEETTPYEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEET 1070
 QY 702 KKP-----AKPE---LAFT--TKGAPSTSDKAPAPTTKAPAPTTKAPAPTTKAP 750
 DB 1071 YEPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAP 1130
 QY 751 PTPTEPPTTSEVSTPTTKEPT--TLHKSDESTPELSAPPKA-LENSPKPEG--- 804
 DB 1131 EETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTE 858
 QY 805 --VPTTTPAPATKPEMTTAKDKTTERDLRTT---ETTTAPKMTKETAATTEKTES 1189
 DB 1190 TVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEET 1244
 QY 859 KITATTOVSTTQDT-----PRKIT--LKTTLAPKVTTKKITTTTTEIM 905
 DB 1245 TTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEET 1304
 QY 906 NKPEETAKPKDRAT--NSKATPKP--OKPTKAPKKPT-----STKAP-KTMPR 949
 DB 1305 PTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTE 1364
 QY 950 VRKPKTTPTRKMTSTMBELNPTSRIAEAMLQTTTP--NOTPNSKLVEVNPSESDAG 1005
 DB 1365 TDEPTDEPTDE--PSDEPTDEPTDEPTDEPTDEPTDEPTDEPTDEPTDEPTDEPTDE 1420

RESULT 9
 Q9SPMO PRELIMINARY; PRT: 1315 AA.
 AC Q9SPMO; 01-MAY-2000 (Tremblrel. 13, created)
 DT 01-MAY-2000 (Tremblrel. 13, last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
 DE EXTENSIN-LIKE PROTEIN.

GN pex2.
 OS Zea mays (Maize).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
 OC Panicoideae; Andropogoneae; Zea.
 RN [1]
 NCBI_TaxID=4577;
 RP SEQUENCE FROM N.A.
 RC TISSUE=POLLEN.
 RA Stratford S., Barnes W., Golbiewski A., Cotter R., McCormick S.,
 RA Hohorst D., Gao M., Showalter A., Hedinger P.A.,
 RT "Polien Extensin-Like (Pex) genes in a Monocot and a Dicot."
 RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF159297; AAD55980.1;
 DR Interpro: IPR001611; LRR.
 DR Interpro: IPR001998; Xylose_1som.
 DR Interpro: IPR002965; P_rich_extensn.
 DR Interpro: IPR003592; LRR_out.
 DR Pfam: PF00560; LRR; 3.
 DR PRINTS: PRO1217; PRICHEXTENS.
 DR PROSITE: PS00172; XYLOSE_1ISOMERASE_1; UNKNOWN_1.
 DR SMART: SM00370; LRR; 3.
 SQ SEQUENCE 1315 AA; 134401 MW; 64C97A2A01F0936F CRC64;

Query Match 11.5%; Score 811.5; DB 10; Length 1315;
 Best Local Similarity 24.6%; Pred. No. 2e-46;
 Matches 262; Conservative 89; Mismatches 473; Indels 243; Gaps 35;

QY 73 RCFESFGRGECODDQCKRYDCCEPESFCALVNDKNTKKKPTP-----K 122
 DB 390 QCAPVLARPEVVCCKH-----VCAGY-----PRGCGPSSSVGK 425
 QY 123 PLYVDEAGSLDNGDKRVTTPTDSTTOHNKVSSTPKITAKPINRPSLP----- 173
 DB 426 PPSVPG-----KPAAPAPMTPIPIPPDVSP-----EPL-PEPSVVPAPAPKMT 469
 QY 174 -NSDTSKETSLSYVKKETTYETKTTTINKQTSIDGKEKTSKAKETOSIEKTSAKDLAPTS 232
 DB 470 LRSPAPDEVTPT-----PVPKASPPGTSPPASGAPLQAPPAASPPATPVKSSPPA 525
 QY 233 KV-----LAKPTKAEETTTGSPALTTKPEPTTPKEPASTPKPEPTTTIKSAPTTK 278
 DB 526 AVLLPPAPKTPSPAPVAVASPPPEAVSSPQOVKSSPPAPVAVASPPPKSP--PPAPRV 583
 QY 279 KSAPTTPKAPAPTTKSAPTTP--KEPAP-----TTKAPAPTTKAPAPTTKAPAPTT 330
 DB 584 ASPPLMKSPPPAPVAVASPPQPLKSPPEVLMVLTSPVSKSPPPVAVASPPPVKSSPPL 643
 QY 331 TTKSAPTTPKPE-----PAP--TTP-----KKAPATTPKAPAPTTKPE--PTTPPKAPAPTTK 379
 DB 644 AVSSSPSPVVKLPPLAPAKSTTPPEEEKPIPTPVKSSPPEKSLPPTTLTSPPOEK 703
 QY 380 EPAPTTKAPAPTAAPKAPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAPTTK 439
 DB 704 PIPSPSPKSPPPSPPEVETLPPSKSSPPEVSSPQAPKSSP-----PAPVS--SPP 756
 QY 440 TTKEPAPTTKSAPTTPKPEPTTTKAPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAPTT 499
 DB 757 LKSSPPPVSPSPPTPKSSPPLAVSSSPPOVEKTSPPAPVAVASPPPKSSPPLAVSS 816
 QY 500 APTTTKAPAPTAAPKAPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAPTTK 559
 DB 817 PPQVEKTSPPAPVAVSSPPPTPKSSPPLAVSSSPPOVEKTSPPAPVAVSSPPEKSPSS 876
 QY 560 PPEPTTPTEPAPPTTPKAAAPNTPKAPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAPTT 619
 DB 877 SVSSPPTPVKSSPPAPLSSPPMTPKSSSPPAHVSSPPEAKSSPPLAPLSSPSEKSP 936
 QY 620 A-----PTTLKAPAPTTKAPAPKELATTTTKEPTSTSDKAPAPTT-----TPKG 663
 DB 937 SSPPVKETSPPATVSSPPTPKSSPP-----APVSSPPPVAVKSSPPAPVAVSSPPPTPK 993

DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
 DE Y51B1A.1 PROTEIN.
 GN Y51B1A.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Poloderinae; Caenorhabditis.
 ON NCBI_TaxID=6239;
 RX NCBI_TaxID=6239;
 RC SEQUENCE FROM N.A.
 RP STRAIN-BRISTOL N2;
 RX MEDLINE=9069613; PubMed=9851916;
 RA None;
 RT "Genome sequence of the nematode C. elegans: a platform for
 RT investigating biology of the C. elegans Sequencing Consortium."
 RT Science 282:2012-2018(1998).
 RN [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Cotton M.;
 RT "The sequence of C. elegans cosmid Y51B1A."
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RN SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC006797; AAF60743.1;
 DR InterPro; IPR002965; P-rich_extensn.
 DR PRINTS; P01217; PRICHEXTENS.
 SQ SEQUENCE 1079 AA; 110532 MW; 8DBDE3824CF80CA1 CRC64;

Query Match 13.8%; Score 971.5; DB 5; Length 1079;

Best Local Similarity 29.5%; Pred. No. 2.9e-57;
 Matches 330; Conservative 129; Mismatches 481; Indels 177; Gaps 44;

QY 106 EVKDNKNNKTKKPPKPPVVDAGSGLDNGDFKV-----TTPDDSTTQ-HNNVSTSP 157
 DB 31 EIADCRANOTPHMLP-----STLSVDMETPSLVLSTSTSSTPIKETTTTAP 81
 QY 158 KITAKPIMPRLPSLPSNDSKETSIVNKETT-VEKKEETTTNKQSTIDGKEKTTSAKE 216
 DB 82 ETTSTEP-----PSSSTTPVQTTTITAPETTSSTEPSSSTTPVQTT-----TTTAP 128
 QY 217 TOSIEKTSKADLAPTSKVLAKPTPKAETT-TKGP-ALTPPKPEPTPTPKPEASTPKET 274
 DB 129 TSTEPSSS---TSPVQTTTITAPETTSSTEPSSSTTPVQTTTITAPETTSSTEPSSS 184
 QY 275 PTTIKSAPTTPKPAPTTTKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKS 334
 DB 185 TSPVQTTTITAPETTSSTEPSSSTTPVQTTTITAPETTSSTEPSSSTTPVQTTTITAP 240
 QY 335 APPTPKPEAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPK 394
 DB 241 APETTSSTEPSSSTTPVQTTTITAPETTSSTEPSSSTTPVQTTTITAPETTSSTEPSSS 299
 QY 395 KKPAPTPKPAPTTPKPE-----APTTKPSPTTPKEPAPTTPKPAPTTPK 440
 DB 300 TTPVQTTTITAPETTSSTEPSSSTTPVQTTTITAPETTSSTEPSSSTTPVQTTTITAP 359
 QY 441 TK--EPAPTTKSAPTTPKPEPTTPKPE-APTPKPEPAPTTPKPAPTTPKPAPTTPK 497
 DB 360 TREPPSSSTTPVQTTTITAPETTSSTEPSSSTTPVQTTTITAPETTSSTEPSSST 415
 QY 498 EPAPTTKPKPAPTPKPE-APTPKETAATTPKPKLLPTTPKELAPITTPKPAPTTPKPEEL 555
 DB 416 TTPVQTTTITAPETTSSTEPSSSTTPVQTTTITAPETTSSTEPSSSTTPVQTTTITAP 471
 QY 556 APPTPEPTPTTPKEPAPTTPKKAAPTTPKPE-APTPKPEPAPTTPKPAPTTPKPAPTTPK 613
 DB 472 PETTSTEP-PSSSTTPVQTTTITAPETTSSTEPSSSTTPVQTTTITAPETTSSTEPSSS 529
 QY 614 TTPGTAATTTKEPAPTTPKPAKPAKELAPITTPKEPTSTSDKP-APTPKGTATPTPKPE 672

DB 530 TTPVQTTTITAPETTSSTEPSSSTTPVQTTTITAPETTSSTEPSSSTTPVQTTTITAP 588
 QY 673 APPTPKPE-APTPKGTATTTKEPAPTTPKPAKPAKELAPITTPKGTSTSDKP-APT 729
 DB 569 --TSTEPSSSTTPVQTTTITAPETTSSTEPSSSTTPVQTTTITAPETTSSTEPSSST 646
 QY 730 KETAPPTPKPAPTTPK-----PAPTPEPTTPPTS-----EVSPTTPKEPTTI 776
 DB 647 PVQTTTITAPETTSSTEPSSSTTPVQTTTITAPETTSSTEPSSSTTPVQTTTITAPETS 706
 QY 777 HKSPESPPELSAEP--PKALENSPKPEGVPTTKTPAATPEMTTAKDKTERDLRT 834
 DB 707 TEPPSSSTTPVQTTTITAPETTSSTEPSSSTTPVQTTTITAPETTSSTEPSSSTTPVQTT 766
 QY 835 ---PETTAAPKMKETAATTEKTESKINATTOVSTSTODPTPKITTTTITAP 890
 DB 767 TTTAPETTSSTEPSSSTTPVQTTTITAPETTSSTEPSSSTTPVQTTTITAPETTSSTEP 824
 QY 891 KYTTTKKTTTTEINMKPEETA-KPKDRATNSKATTPPKPKPKPAKPKPTSKPKMPR 949
 DB 825 PSSSNTPVQTTTITAPETTSSTEPSSSTTPVQTTTITAPETTSSTEPSSSTTPVQTTT 882
 QY 950 VRKPKTTPPKMTSTMPELNPTSRLEAMLOTTTRPNQTPSKLVEVNPKEADGAG 1009
 DB 883 ITAPETTSSTEPSSSTTPVQTTTITAPET--TSTEP--PSSSTTPVQTTTITAPETTS 936
 QY 1010 -ETPHMLLRHYMPEVTP--DMOYLPRV-----PNOGIINPMLSDEFN---- 1051
 DB 937 TEPPSSSTTPVQTTTITAPEDCSLSYIDRVYVFTTEWEMENKRDILIQSYDSPRTAF 996
 QY 1052 -----ICNGKPVGGLTLRNGTLVAFRGHYWMLS-----PSPSPPAR 1091
 DB 997 VSTEDICTATALLCIATYSGGISNL-NATL-----FTGSDGSSIDLPE----- 1040
 QY 1092 ITEWGLPSPIDVYTRCNCEGKTFEPKQKQYRFTN 1128
 DB 1041 ---YNPGLGLE-TPMEINCEGKN-----WSYNN 1064

RESULT 6
 ID Q917S1 PRELIMINARY; PRT; 1049 AA.
 AC Q917S1;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
 DE CG5228 PROTEIN.
 GN CG5228.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 ON NCBI_TaxID=7227;
 RX [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN-BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amaralides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.C., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Bolashkov S.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolashkov S.,
 RA Borjesson D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

[illegible]

Db	1024	OYXNIDVPTRTARALTTRSCGLSKRWNC	1054
RESULT	4		
ID	077765	PRELIMINARY;	PRT; 401 AA.
AC	077765;		
DT	01-NOV-1998	(TREMBLrel. 08, Created)	
DT	01-NOV-1998	(TREMBLrel. 08, Last sequence update)	
DT	01-JUN-2001	(TREMBLrel. 17, Last annotation update)	
DE	SUPERFICIAL ZONE PROTEIN (FRAGMENT).		
OS	Bos taurus (Bovine).		
OC	Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;		
OC	Bovidae; Bovinae; Bos.		
OX	NCBI_TaxID=9913;		
RM	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=ARTICULAR CARTILAGE;		
RA	Schumacher B.L., Hughes C.E., Kuettner K.E., Caterson B.,		
RA	Angelote M.B.;		
RT	"Immunodetection and Partial cDNA Sequence of the Proteoglycan,		
RT	Joint's",		
RL	J. Orthop. Res. 0:0-0(1998).		
DR	EMBL; AF056218; AAD13404.1;		
DR	InterPro; IPR000585; Hemopexin.		
DR	Pfam; PF00045; hemopexin. 2.		
DR	PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.		
DR	SMART; SM00120; HX; 2.		
FT	NON_TER		
SO	1		
SEQUENCE	401 AA: 44952 MW; 86147CG9AFBB73D7 CRC64;		

Query Match	24.3%	Score 1713.5	DB 6	Length 401
Best Local Similarity	80.2%	Pred. No. 1,46-107		
Matches 320	Conservative 24	Mismatches 52	Indels 3	Gaps 3
QY 914 PKDRATNSKATYPPKPPQPKPTKAPKPKPTSTKTKPKPTMVRVKKPKTTTPPKK-MTSTMPELNPT 972				
Db 5 PKGRATNSQVTPPKPKQPKTKKPKKPTSTKKPRTP-PRVKKPKPTTPPKTKTTSAMPEPTPT 63				
QY 973 SRTAEMLQTTTRPNOTPNSKILVEVNPKSEDAGAEGETPHMLRLPHVEMPEVTPDMQVL 1032				
Db 64 S-LPEMLQTTTRPPTPPTNSEIIDVNSNEDGDAAEKPHMIFRPVLTPIVIGETETI 122				
QY 1033 PRYPNGIILINMLSDENINCNKGPVSDGLTTLRNGTLVAFRGHYFWMLSPFSPSPARRI 1092				
Db 123 VNGPSQSGEGINPMFSDENILNCGRVGLATLLRNGTLVAFRGHYFWMLPPTPPPPRRRI 182				
QY 1093 TEVMGIPSPIDIVETFRNCCEKGTPEFPKDSQWPRPTNDIKAGIGPKIIFGFGGLTGQIVA 1152				
Db 183 TEVMGIPSPIDIVETFRNCCEKGTTEFFGSGQWRPTNDIKAGIGPKIISGFGGLNKQIVA 242				
QY 1153 AASTAKKYNMPSVYFEFRGGSIQOYLYKQEPVQKCPGRPALNYPVYEGMTQVRRRRE 1212				
Db 243 AASTAQKYSRPSVYFEFRGGSVQOYLYKQEPQKCTGRPALNYSVYGETAQVRRRRE 302				
QY 1213 RAIGSQVHTIRIYSPARLAYQDKGYLHNEVYSLMLGCLPNVYTSALSPINIKKPDGY 1272				
Db 303 RAIGSQVHTIRIYTPVPRVYQDKGLHNEVYSLMLGCLPNVYTSALSPINIKKPGY 362				
QY 1273 DYVAESKQDYINIDVPSRTARATYTRSSGQTLISVWVWVNC 1211				
Db 363 DYVALSKQDYINIDVPSRTARATYTRSSGQTLISVWVWVNC 401				
RESULT 5				
Q9N4S7	PRELIMINARY	PRT: 1079 AA.		
Q9N4S7				
Q9N4S7				
01-OCT-2000 (T-Emblrel. 15, Created)				
01-OCT-2000 (T-Emblrel. 15, Last sequence update)				

```

QY 107 -----VKDNKKNRKTKKPKPPVVDGAGSLDNGDKVTPDST 147
Db 181 KIKSSKSNANRELQKKLKVKNKKNRKTKKPKPPVVDGAGSLDNGDKVTPDST 240
QY 148 TOHNKVSISPKITTAKPINPRPSLPPNSDTSKETSLTVNKEETTVETKETTNNKOTSDG 207
Db 241 TOHNKVSISPKITTAKPINPRPSLPPNSDTSKETSLTVNKEETTVETKETTNNKOTSDG 300
QY 208 KEKTSKAKETOSIEKTSAKDLAPTSKVLAKPPTKAEITTKGALTPPKBPTTPKEPAS 267
Db 301 KEKTSKAKETOSIEKTSAKDLAPTSKVLAKPPTKAEITTKGALTPPKBPTTPKEPAS 360
QY 268 TTPKEPPTTIKSAPTTPKEBAPTTPKSAPTTPKEBAPTTPKSAPTTPKSAPTTPK 327
Db 361 TTPKEPPTTIKSAPTTPKEBAPTTPKSAPTTPKEBAPTTPKSAPTTPKSAPTTPK 420
QY 328 APPTTKSAPTTPKEBAPTTPKPKAPPTTPKKEBAPTTPKKEBAPTTPKKEBAPTTPK 387
Db 421 APPTTKSAPTTPKEBAPTTPKPKAPPTTPKKEBAPTTPKKEBAPTTPKKEBAPTTPK 480
QY 388 EPAPJTAKKAPATTPKEBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPT 447
Db 481 EPAPJTAKKAPATTPKEBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPT 540
QY 448 TTKSAPTTPKSPSTTPKEBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPT 507
Db 541 TTKSAPTTPKSPSTTPKEBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPT 600
QY 508 APAPKEBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPTTPK 567
Db 601 APPTTKKEBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPTTP 660
QY 568 PEAPAPPTTPKKAAPPTTPKEBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPT 627
Db 661 PEAPAPPTTPKKAAPPTTPKEBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPT 720
QY 628 APPTTKKAPAKELAPPTTKKEPTSTSDKRAPPTPKGTAAPTPKKEBAPTTPKKEBAPTTPK 687
Db 721 APPTTKKAPAKELAPPTTKKEPTSTSDKRAPPTPKGTAAPTPKKEBAPTTPKKEBAPTTPK 780
QY 688 TAPPTLKAPAPPTTPKPKAPAKELAPPTTKGPTSTSDKRAPPTPKGTAAPTPKKEBAPTTPK 747
Db 781 TAPPTLKAPAPPTTPKPKAPAKELAPPTTKGPTSTSDKRAPPTPKGTAAPTPKKEBAPTTPK 840
QY 748 KAPAPPTTPPPTSEVSTPTTKKEPTTIHKSPPDESTBELSAEPTPKALENSPKBGPVT 807
Db 841 KAPAPPTTPPPTSEVSTPTTKKEPTTIHKSPPDESTBELSAEPTPKALENSPKBGPVT 900
QY 808 TKTPAATKPEMTTAKDKTTERDLKTTTETTTAAKMKKETATTTTEKTESKITATTOY 867
Db 901 TKTPAATKPEMTTAKDKTTERDLKTTTETTTAAKMKKETATTTTEKTESKITATTOY 960
QY 961 TKTPTODTTPPKITTLKTTTLAPKVTTKKITTTEINMKPEELAPKDRATNSKATTPK 927
Db 961 TKTPTODTTPPKITTLKTTTLAPKVTTKKITTTEINMKPEELAPKDRATNSKATTPK 1020
QY 928 POKPTAKPKPTSTTKPKKPTMPKPKPTTPPKKATSTYMBELNPSRIEAMLOTTTRN 987
Db 1021 POKPTAKPKPTSTTKPKKPTMPKPKPTTPPKKATSTYMBELNPSRIEAMLOTTTRN 1080
QY 988 QTPRSKIVENPNSSEDAAGEGEPHMLRPHAFMEVTPDDMDYLRVPNOGIINPMLS 1047
Db 1081 QTPRSKIVENPNSSEDAAGEGEPHMLRPHAFMEVTPDDMDYLRVPNOGIINPMLS 1140
QY 1048 DETNICKGKPVGDLTTLRNGTILVAFRGHYTWMLSPPSPSPARITIEVNCISPIDVFT 1107
Db 1141 DETNICKGKPVGDLTTLRNGTILVAFRGHYTWMLSPPSPSPARITIEVNCISPIDVFT 1200
QY 1108 RCNCEGKTFPFKDSQVWRFNDIKDAGYPRPIFGFGGLGGOIYALALSTAKYNNWESVY 1167
Db 1201 RCNCEGKTFPFKDSQVWRFNDIKDAGYPRPIFGFGGLGGOIYALALSTAKYNNWESVY 1260
QY 1168 FFKRGSGIOQYIYKQEBVQKCGRRPALNYPVYGEMTVORRRRPERAIGPSOHTIRIQY 1227

```

```

Db 1261 FFKRGSGIOQYIYKQEBVQKCGRRPALNYPVYGEMTVORRRRPERAIGPSOHTIRIQY 1320
QY 1228 SPATLADODGVILHNEKVSILHMGLEPNVTSAILSPNFKRPDGYDYAFASKDQYNTIDY 1287
Db 1321 SPATLADODGVILHNEKVSILHMGLEPNVTSAILSPNFKRPDGYDYAFASKDQYNTIDY 1380
QY 1288 PSRTARATITRSQOTLSKWYNCP 1311
Db 1381 PSRTARATITRSQOTLSKWYNCP 1404

RESULT 3
QJUM99 PRELIMINARY; PRT: 1054 AA.
QJUM99:
AC 01-OCT-2000 (TREMBLERel. 15, Created)
DT 01-OCT-2000 (TREMBLERel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLERel. 17, Last annotation update)
DE MRNA, COMPLETE CDS, SIMILAR TO MEGAKARYOCYTE STIMULATING FACTOR
DE PRECURSOR AND CARTILAGE SOBERFICIAL ZONE PROTEIN.
GN PRC4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN (1)
RP SEQUENCE FROM N.A.
RA Ikegawa S., Nakamura Y.;
RT "A novel mouse gene highly homologous to a human gene encoding
RT megakaryocyte stimulating factor precursor and cartilage superficial
RT zone protein."
RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: AB034730; BAA92310.1;
DR MGD: MGI:1891344; Prg4.
DR InterPro: IPR000365; Hemopekin.
DR InterPro: IPR002965; P_fich_extensin.
DR InterPro: IPR001212; Somatomedin_B.
DR Pfam: PF01033; Somatomedin_B; 2.
DR PRINTS: PR01217; PRICHEXTENSIN.
DR PRINTS: PR00022; SOMATOMEDINB.
DR SMART: SM00120; HX; 2.
DR SMART: SM00201; SO; 2.
DR PROSITE: PS00524; HEMOPEKIN; UNKNOWN_1.
DR PROSITE: PS00524; SOMATOMEDIN_B; 2.
SQ SEQUENCE 1054 AA; 115991 MW; 4FC64BFA42283235 CRC64;

Query Match 51.5%; Score 3640.5; DB 11; Length 1054;
Best Local Similarity 52.4%; Pred. No. 1,2e-236;
Matches 740; Conservative 55; Mismatches 159; Indels 457; Gaps 22;

QY 1 MAWKLTPYLLILSVFVIOVSSQDILSCAGRCGEGYSRDATCNDYNCQHYMECCPP 60
Db 1 MGWKLTPYLLILSVFVIOVSSQDILSCAGRCGEGYSRDATCNDYNCQHYMECCPP 60
QY 61 KRVTAEISCKGRCEPESREBERECDDAOCKKXKCPDYESFCAB----- 106
Db 61 KRVTAEISCKGRCEPESREBERECDDAOCKKXKCPDYESFCAB----- 106
QY 61 KRVTAEISCKGRCEPESREBERECDDAOCKKXKCPDYESFCAB----- 106
Db 61 KRVTAEISCKGRCEPESREBERECDDAOCKKXKCPDYESFCAB----- 106
QY 107 -----VKDNKKNRKTKKPKPPVVDGAGSLDNGDKVTPDST 147
Db 107 -----VKDNKKNRKTKKPKPPVVDGAGSLDNGDKVTPDST 147
QY 107 -----VKDNKKNRKTKKPKPPVVDGAGSLDNGDKVTPDST 147
Db 107 -----VKDNKKNRKTKKPKPPVVDGAGSLDNGDKVTPDST 147
QY 181 SKNSANRELQKKLKVKNKKNRKTKKPKPPVVDGAGSLDNGDKVTPDST 240
Db 181 SKNSANRELQKKLKVKNKKNRKTKKPKPPVVDGAGSLDNGDKVTPDST 240
QY 151 NKVSTSPKTTAKKPINPRPSLPPNSDTSKETSLTVNKEETTVETKETTNNKOTSDG 210
Db 241 SKVATSPKTTAKKPINPRPSLPPNSDTSKETSLTVNKEETTVETKETTNNKOTSDG 299
QY 211 TTSKAKETOSIEKTSAKDLAPTSKVLAKPPTKAEITTKGALTPPKBPTTPKEPAS 270

```

DR PRINTS; PRO0022; SOMATOMEDINB.
 DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
 DR PROSITE; PS00534; SOMATOMEDIN_B; 2.
 DR SMART; SM00120; HX; 2.
 DR SMART; SM00201; SO; 2.
 SO SEQUENCE 1404 AA; 151090 MW; AABD7AD19B35F4F6 CRC64;

Query Match 99.2%; Score 7007.5; DB 4; Length 1404;
 Best Local Similarity 93.4%; Pred. No. 0;
 Matches 1311; Conservative 0; Mismatches 0; Indels 93; Gaps 1;

1 MAMKLPYILLLLSVFYIQOVSSQDLSSCAGRCGEGSRDATNCNDYNOCHMECCPDF 60
 1 MAMKLPYILLLLSVFYIQOVSSQDLSSCAGRCGEGSRDATNCNDYNOCHMECCPDF 60
 61 KRVCTAELSCGRCFESFERGECDDAOCCKYDKCCPDYESFCAE----- 106
 61 KRVCTAELSCGRCFESFERGECDDAOCCKYDKCCPDYESFCAE----- 106
 107 ----- 106
 121 PPSGASQTIKSTTKRSRPPNKKTKKVIIESEITEEHSVENQESSSSSSSSSTIW 180
 107 -----VKDNKKNRTKKKPKPKPVYDEAGSLDNGDFKVTTPDIST 147
 181 KIKSSKSAANRELQKLVKDKNKKTKKPKPKPVYDEAGSLDNGDFKVTTPDIST 240
 148 TONHKVSTPKITAKPINRPSLPPNSDTSKETSJLVNKETTVEKETTNNKOTSDG 207
 241 TONHKVSTPKITAKPINRPSLPPNSDTSKETSJLVNKETTVEKETTNNKOTSDG 300
 208 KEKTSKAKETQSIKTSANDLAFTSKVLAKPTPKAETTKGPAITTPKEPTTPKEBAS 267
 301 KEKTSKAKETQSIKTSANDLAFTSKVLAKPTPKAETTKGPAITTPKEPTTPKEBAS 360
 268 TTPKEPTTPKISAPTTKEBPATTTKSAPTTPKKEBPATTTKEBPATTTKEP 327
 361 TTPKEPTTPKISAPTTKEBPATTTKSAPTTPKKEBPATTTKEBPATTTKEP 420
 328 APPTTKSAPTTPKEBPATTTKRAPTTPKKEBPATTTKEPTTPKEBPATTTK 387
 421 APPTTKSAPTTPKEBPATTTKRAPTTPKKEBPATTTKEPTTPKEBPATTTK 480
 388 EPAPTPAKKAPATTPKEBPATTTKRAPTTPKKEBPATTTKSAPTTPKEBPAT 447
 481 EPAPTPAKKAPATTPKEBPATTTKRAPTTPKKEBPATTTKSAPTTPKEBPAT 540
 448 TTKSAPTTPKESPTTPKEBPATTTKRAPTTPKKEBPATTTKSAPTTPKEBPAT 507
 541 TTKSAPTTPKESPTTPKEBPATTTKRAPTTPKKEBPATTTKSAPTTPKEBPAT 600
 508 APATKKEBPATTPKEBPATTTKRAPTTPKKEBPATTTKEPTTPKEBPATTT 567
 601 APATKKEBPATTPKEBPATTTKRAPTTPKKEBPATTTKEPTTPKEBPATTT 660
 568 PEEBPATTPKAAANPTTPKEBPATTTKRAPTTPKKEBPATTTKEPTTPKEBPAT 627
 661 PEEBPATTPKAAANPTTPKEBPATTTKRAPTTPKKEBPATTTKEPTTPKEBPAT 720
 628 APPTPKKAPAPKELAPTTKEPTTSDKRAPTPPKGTAPTTKEBPATTPKKEBPATTP 687
 721 APPTPKKAPAPKELAPTTKEPTTSDKRAPTPPKGTAPTTKEBPATTPKKEBPATTP 780
 688 TAPPTLKEBPATTPKRAPELAPTTKGTSTSDKRAPTPPKGTAPTTKEBPATTPK 747
 781 TAPPTLKEBPATTPKRAPELAPTTKGTSTSDKRAPTPPKGTAPTTKEBPATTPK 840
 748 KRAPTPPETPTTSEVSTPTTKEPTTIHKSPTDESTPELSAEPKALESPKEGVP 807
 841 KRAPTPPETPTTSEVSTPTTKEPTTIHKSPTDESTPELSAEPKALESPKEGVP 900
 808 TKTPAATKPEMTTAAKDTTERDLRTTPETTTAAKMKKETATTEKTSKITATTTGV 867

Db 901 TKTPAATKPEMTTAAKDTTERDLRTTPETTTAAKMKKETATTEKTSKITATTTGV 960
 QY 868 TSTTQDTPPFKITTLTKTTTLAKVYTTTKKITTITTEINKEPEFAKPKDRATNSKATTPK 927
 Db 961 TSTTQDTPPFKITTLTKTTTLAKVYTTTKKITTITTEINKEPEFAKPKDRATNSKATTPK 1020
 QY 928 POKPKPAKPKPTSTPKPTMPVRKPKTTPPRKMTSTMPELNPTSRIAEMIQTTTRPN 987
 Db 1021 POKPKPAKPKPTSTPKPTMPVRKPKTTPPRKMTSTMPELNPTSRIAEMIQTTTRPN 1080
 QY 988 QTPNSKLEVPNPKSPDAGAGETPHMLLRPHVMPREVTDMQDLPVRPNOGIIINPMLS 1047
 Db 1081 QTPNSKLEVPNPKSPDAGAGETPHMLLRPHVMPREVTDMQDLPVRPNOGIIINPMLS 1140
 QY 1048 DETNLCNKPVDGLTTLRNGTLVAFRGHYFWMLSPEFSPSPARRITEWNGISPDIYVT 1107
 Db 1141 DETNLCNKPVDGLTTLRNGTLVAFRGHYFWMLSPEFSPSPARRITEWNGISPDIYVT 1200
 QY 1108 RCNCEGKTFEFKDSQYRFTNDIKDAGYRPIKRGGLTGQIVAAIATAKYNMPESY 1167
 Db 1201 RCNCEGKTFEFKDSQYRFTNDIKDAGYRPIKRGGLTGQIVAAIATAKYNMPESY 1260
 QY 1168 FFKRGSSIQOYITKOEPEVOKCPGRPALNYPVYGEOTVRRRFRERAIQSPOTHTIRIQ 1227
 Db 1261 FFKRGSSIQOYITKOEPEVOKCPGRPALNYPVYGEOTVRRRFRERAIQSPOTHTIRIQ 1320
 QY 1228 SPARLAVQDKGVLANHEKVSILMRGLPNVTSAISLPNIRAKDGYDYAFSKDOYNNIDV 1287
 Db 1321 SPARLAVQDKGVLANHEKVSILMRGLPNVTSAISLPNIRAKDGYDYAFSKDOYNNIDV 1380
 QY 1288 PSRTARAITTRSGQITLSKWMYNC 1311
 Db 1381 PSRTARAITTRSGQITLSKWMYNC 1404

RESULT 2
 ID Q9BX49 PRELIMINARY: PRT; 1404 AA.
 AC Q9BX49:
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DE BG174L6.2 (MSF: MEGAKARYOCYTE STIMULATING FACTOR).
 GN BG174L6.2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Wray P.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL133553; CAC36090.1;
 SO SEQUENCE 1404 AA; 151076 MW; 782A11746B3FDEE5 CRC64;

Query Match 99.0%; Score 6992.5; DB 4; Length 1404;
 Best Local Similarity 93.2%; Pred. No. 0;
 Matches 1308; Conservative 0; Mismatches 3; Indels 93; Gaps 1;

1 MAMKLPYILLLLSVFYIQOVSSQDLSSCAGRCGEGSRDATNCNDYNOCHMECCPDF 60
 1 MAMKLPYILLLLSVFYIQOVSSQDLSSCAGRCGEGSRDATNCNDYNOCHMECCPDF 60
 61 KRVCTAELSCGRCFESFERGECDDAOCCKYDKCCPDYESFCAE----- 106
 61 KRVCTAELSCGRCFESFERGECDDAOCCKYDKCCPDYESFCAE----- 106
 107 ----- 106
 121 PPSGASQTIKSTTKRSRPPNKKTKKVIIESEITEEHSVENQESSSSSSSSSTIR 180

Mon Apr 29 08:35:26 2002

aar.rspt

Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:30:16 ; Search time 210.7 Seconds
(without alignments)
910.124 Million cell updates/sec

Title: AAR
Perfect score: 7064
Sequence: 1 MAMKTLPIYLILLSEVIO.....ARAITSNGQTLKWNCP 1311

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 segs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP:archaea:***
2: SP:bacteria:***
3: SP:fungi:***
4: SP:human:***
5: SP_invertebrate:***
6: SP_mammal:***
7: SP_mhc:***
8: SP_organelle:***
9: SP_phage:***
10: SP_plant:***
11: SP_rodent:***
12: SP_virus:***
13: SP_vertibrate:***
14: SP_unclassified:***

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7007.5	99.2	1404	4 Q92954	Q92954 homo sapien
2	6992.5	99.0	1404	4 Q9BX49	Q9BX49 homo sapien
3	3640.5	51.5	1054	11 Q9JN99	Q9JN99 mus musculu
4	1713.5	24.3	401	6 Q77765	Q77765 bos taurus
5	971.5	13.8	1079	5 Q9N457	Q9N457 caenorhabd
6	936.5	13.3	1049	5 Q9I761	Q9I761 drosophila
7	865.5	12.3	1795	5 Q96449	Q96449 drosophila
8	855	12.1	1489	10 Q96449	Q96449 phytothor
9	811.5	11.5	1315	10 Q9SPM0	Q9SPM0 zea mays (m
10	802	11.4	1274	5 Q20007	Q20007 caenorhabd
11	800.5	11.3	1480	10 Q9L1F8	Q9L1F8 arabidopsis
12	784	11.1	1188	10 Q41805	Q41805 zea mays (m
13	776.5	11.0	2187	11 P70670	P70670 mus musculu
14	752	10.6	2112	5 Q9VEL9	Q9VEL9 mycobacteri
15	715	10.1	763	2 Q9XDH2	Q9XDH2 chlamydomon
16	707.5	10.0	555	2 Q9P6P6	Q9P6P6 gallus gall
17	675.5	9.6	2284	5 Q9YPC1	Q9YPC1 gallus gall
18	671	9.5	1151	11 Q57580	Q57580 gallus gall
19	667.5	9.4	4880	13 Q9JL11	Q9JL11 ratius norv

20	667.5	9.4	5085	11 Q9UK66	Q9UK66 ratius norv
21	667	9.4	6677	5 Q9N435	Q9N435 caenorhabd
22	662	9.4	1229	5 Q94185	Q94185 caenorhabd
23	661	9.4	3507	5 Q23587	Q23587 caenorhabd
24	654.5	9.3	2089	4 Q14676	Q14676 homo sapien
25	653.5	9.3	1514	5 Q9GUW7	Q9GUW7 leishmania
26	651	9.2	1612	5 Q9YX02	Q9YX02 drosophila
27	649	9.2	2768	5 Q9YX02	Q9YX02 drosophila
28	647	9.2	4833	11 Q90YX6	Q90YX6 mus musculu
29	647	9.2	5038	11 Q90YX7	Q90YX7 mus musculu
30	646.5	9.2	3570	4 Q99552	Q99552 homo sapien
31	635.5	9.0	7962	4 Q10465	Q10465 homo sapien
32	633	8.9	489	10 Q41707	Q41707 vigna ungu
33	632	8.9	761	10 Q9X010	Q9X010 arabidopsis
34	632	8.9	6632	5 Q17362	Q17362 caenorhabd
35	630.5	8.9	990	13 Q91803	Q91803 xenopus lae
36	626.5	8.8	971	5 Q9XVS4	Q9XVS4 caenorhabd
37	625	8.8	6642	5 Q01761	Q01761 caenorhabd
38	622.5	8.7	839	2 Q9RX57	Q9RX57 deinoceus
39	615	8.6	2344	5 Q9N3Y8	Q9N3Y8 caenorhabd
40	608	8.6	801	5 Q23635	Q23635 caenorhabd
41	607.5	8.6	924	12 Q99307	Q99307 epstein-bar
42	605	8.6	379	5 Q27929	Q27929 drosophila
43	600.5	8.5	1893	5 Q9NKM1	Q9NKM1 drosophila
44	598.5	8.5	409	10 Q9SBM1	Q9SBM1 volvox cart
45	577	8.2	956	10 Q9LJ64	Q9LJ64 arabidopsis

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	1404 AA.
ID	Q92954			
AC	Q92954:			
DT	01-FEB-1997 (TREMBLrel. 02, Created)			
DT	01-FEB-1997 (TREMBLrel. 02, last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, last annotation update)			
DE	MEGAKARYOCYTE STIMULATING FACTOR.			
OS	Homo sapiens (human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,			
RA	Kelleher K., Giannotti J., Calveti J., Fitzgerald M., Kriz M.J.,			
RA	Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,			
RA	Bhatia S., Kriz R., Hewick R., Clark S.C.,			
RL	Blood 78:279-279(1991).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Merberg D.M., Fitz L.J., Temple P., Giannotti J., Murtha P.,			
RA	Fitzgerald M., Scalapeto J., Kelleher K., Preissner K., Kriz R.,			
RA	Jacobs K., Turner K.,			
RL	(in) Preissner K.T., Rosenblatt S., Kost C., Wegerhoff J.,			
RL	Mosher D.F. (eds.);			
RL	Biology of vitronectins and their receptors., pp.45-52,			
RL	Elsevier Science Publishers B.V. (1993).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RA	Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,			
RA	Kelleher K., Giannotti J., Calveti J., Fitzgerald M., Kriz M.J.,			
RA	Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,			
RA	Bhatia S., Kriz R., Hewick R., Clark S.C.,			
RL	Submitted (SEP-1996) to the EMBL/Genbank/DBJ databases.			
RL	EMBL: U0136; AAB09089.1; -			
DR	InterPro: IPR000585; Hemopexin.			
DR	InterPro: IPR001212; Somatomedin_B.			
DR	InterPro: IPR002400; GF_cysnot.			
DR	Pfam: PF00045; hemopexin; 2.			
DR	Pfam: PF01033; Somatomedin_B; 2.			
DR	PRINTS: PR00438; GFCYSNOT.			

PD 4/23 1999

p. 1-3

```

Db 73 vpsis-ntldtlbpyqlllllslsvnfcrkksqlyk--nfeekfnlasvqssnatqge 129
QY 128 KTTSAKETSIEKTSAKDLAPTSKVLAKETPKAETTTGPAITPKPT-----PTPKEP 183
Db 130 nsngknelev-keessqclqpv-----lpgewltq--aaapqetvclqepvltleep 179
QY 184 ASTPKPEPTPTTIKSAPTTPKAPATTTKSAPTTPKAPATTTKPEAPTTPKAPTTPKAPA 239
Db 180 vltc---qepvltq---qepvltqge--vltqgepvltqge--pvtqgepvltqge 227
QY 240 TTTKEPATTTKSAPTTPKAPATTTKPKAPATTTKPEAPTTPKPEPTPTPKAPATTKEP 299
Db 228 vltqgep---vltqgepvltqge-----pvtqge--pvtqgepvltqge 269
QY 300 A-----PTPKKEPAPTTPKAPATTTKPEAPTTPKAPA-----PTTKPE-----SPTTKPEP 347
Db 270 vltqgepvltleepvltc---qepvltle--pvtqgepvltleepvltqgepvltqge 323
QY 348 APTTTKSAPTTTKAPAPT-----TTKSAPTTPKESPPTTKPEA----- 386
Db 324 ---vltqgepvltqgepvltveehidekkgsegdnlsisiseeteekshlkkksawlkgfr 380
QY 387 -----PTPKKEPAPTTPKAPATTTKPEAPTTPKAPATTTKKPAATAKPEAPTTPK 439
Db 381 gnkndkksknekkpslesvkqnadegqclsdqisvnaqdsvltqgeplaqep-pltqge 439
QY 440 ETAPTTPEKLLPTTPEKLAPTTPEKLAPTTPEELAPTTPEEPTTPEAPTTPEAPTTPKAAP 499
Db 440 lta-----lge-----pltqge-----vltqgeplaqep-----vltqge 470
QY 500 NTPKEPA---PTPKPE---PAPTPEKAPATTPKETAPTTPKGTAPTTTLEKAPATTPK 551
Db 471 lltqgepvltqgeplaqepstlqgeplaqeghadekkasegdnlsisiseeteekshlkkk 530
QY 552 KPAKELAPTTKEPTSTSDKPAATTPKGAATTPPKAPATTPKAPATTPKGTAAPTTL 611
Db 531 kswlkfgrgnknkksknekkpslesvkqnadegqclsdqisvnaqdsvltqgeplaq 590
QY 612 KEPAPTTPKKPAKELAPTTTKGPTSTSDKPAATTPKETAATTPKPEAPTTPKKPAATTP 671
Db 591 qep-pltqgeplaqge--pltqge-----etvteqgeplaqge--vtaqgeplc---qepvt 637
QY 672 PETPPTTSEVSTPTTKKEPTTIHKSPE-----STPELSAEPITPKA----- 713
Db 638 aq-epvltqge---laltqgeplaqeghadekkasegdnlsisiseeteekshlkkkssw 693
QY 714 -----LENSPKPEGPVPTTKTPAATPKEMTTTAKDKTTERDLRTTPE-TTAPAK 761
Db 694 lkfgrgnknkksknekkpslesvkqnadegqclsdqisvnaqdsvltqgeplaqge 753
QY 762 MTKETATTTTEKTESKITATTTQVSTTQDTTPPK--ITTLKTTTTLAPKVTT-----TK 814
Db 754 tlqgelaltqgeplaqge---vltqgeplaqgeplaqgeplaqgeplaqgeplaqge 810
QY 815 KTTTTTEIMMKPEETAPKDRATNSKATTPKPK 848
Db 811 epstlge--hadekkasegdnlsisiseeteek 842

```


XX antimycobacterial; antibiotic; vaccine.
XX
OS Mycobacterium tuberculosis.
XX
PN WO9741252-A2.
XX
PD 06-NOV-1997.
XX
PF 18-APR-1997; 97WO-EP01973.
XX
PR 29-APR-1996; 96DE-4017184.
XX
PA (GBFB) GBF GES BIOTECH FORSCHUNG GMBH.
XX
PI Espitia C, Honisch C, Moreno C, Singh M;
XX
DR WPI: 1997-549750/50.
XX
DR N-PSDB; AAT93610.
XX
PT New DNA and related proteins or RNA derived from M. tuberculosis -
PT used for diagnosis of mycobacterial infections, monitoring
PT vaccination and development of anti-mycobacterial agents
XX
PS Claim 11, Fig 16; 55pp; English.

CC This novel 55 kDa protein is encoded by an open reading frame of
CC a *Mycobacterium tuberculosis* DNA fragment (see AAI993610) containing
CC polymorphic GC-rich sequences. Its amino acid sequence shows
CC a high proline content, but there is no homology to any known
CC proline-rich antigens of mycobacteria. Novel M. tuberculosis
CC proteins (see AAM31851-57) are claimed. These can be produced as
CC recombinant proteins, especially in bacterial, yeast, fungal or
CC higher eukaryote host cells, and used for diagnosing tuberculosis
CC and other mycobacterial infections in humans or animals. The
CC claimed proteins can also be used for epidemiological studies, for
CC monitoring vaccination, and for the development of vaccines and
CC anti-mycobacterial drugs.

SQ Sequence 572 AA;

Query Match	9.38;	Score 612;	DB 18;	Length 572;
Best Local Similarity	30.68;	Pred. No. 5.5e-29;		

Qy	177	PIPIKEBASTPKEERTPTTISAPTTKE-----PAPTTKASAPTTKEAPATTKEBPAP	231
		: : : : : :	
Db	2	paapskskpasopripap---pmpalfmeofpripvpdpdiskeictpapapppriippavp	57
Qy	232	TPPKBEPAPTTTKEBAPTTTKSAPTTTKEBPAP-----TPPKBAPATTPEKEBAPTTTKE	283
Db	58	ipppvpplp-----pvpnklippap-----papvavaavlaagcppllpplpninhppapaa	107
Qy	284	PPPTTPEKEBAPTTKEBAPTTPEKEBAPPAKKBPAP-----TT-----	319
		: : : : : :	
Db	108	pvpgvpylapijushppap--psapvpyvliapijsgrpvsvmkgsticltfcrcvsg	165
Qy	320	-----PKBEA-----PTTKEBAPTTTKEBPTTPEKEBAPTTTKSAPTTKEBAPT	366
		: : : : : :	
Db	166	evlaagalmpsrzsisplttictpalpapl-----pplipplipnlntavppl--pplipv	217
Qy	367	TKSAPTTPE--EBSPTTKEP--APTPEKEBAPTTPKKAPPTPEKEBAPTTPEKEBAPTTTK	423
		: : : : : :	
Db	218	talappplpplapijsgvppapppriippokpwcticpplapappek--tvpv--pppgsccpse	276
Qy	424	KBAPPAPEKEBAPTTPEKETAPTTPKKLPTTPE--KLAPTPEKCAPTTPEELAPTTPEEP	481
		: : : : : :	
Db	277	knpapap--peppekskspalppapapasmgsavrvppspriippap-----paadras	327
Qy	482	TTTTTEBEPAPTTPKA---AAPTPEKEBAPTTPEKEBAPTTPEKEBAPTTTK---ETAPTTPK	533
		: : : : : :	
Db	328	mpaalp--papsppapcrlcpplpssppapap--pappap---ptppkllsanppcpv	379

```

OY      556  GATPITLKEPAPITPKKAPAKELAPITTKETSTTSOKPAFTTKGAPITPKKEPAPTTP 595
Db      380  PPAPn---rppppap---papel-----pappod-----pdpvansppap 416
OY      596  KEPATPTPKGAPITLKEPAPITPKKAPAKELAPITTKGp-----TSTTSOKPAFTTP 648
Db      417  -ppappappapalp-fvnpapppprppaapksrpalpaappappapvratcpppappappap 473
OY      649  KETA--PTTKKEPAP---TTTKKAPITTPETTPPTTSEVSPTTTKKEPTTIHKSPPDSTP 703
Db      474  nsmalppapppdppllacltpappappplpmusppappplppaapdpapppllindpp---sp 530
OY      704  ELSAEPTRKALENSPKREBVGPTTKPATKP 734
Db      531  pla-----pyvgpaplaplplngtrp 549

```

RESULT 15
AAP60570

AC AAP60570;

DT 24-AUG-1991 (first entry)

Sequence of the Falciparum Interspersed Repeat Antigen (FIRA).

KW Malaria vaccine; antigen; epitope.

OS *Plasmodium falciparum*.

PN W08601802-A

PD 27-MAR-1986.

PF 11-SEP-1985; 85WO-0006960

PR 11-SEP-1984; 84AU-0007067

PR 10-SEP-1985; 85AU-0047326

PA (HALL-) HALL INST MED RES.

PI Kemp DJ, Anders R, Coppel RL, Brown G, Saint RB, Cowman AF,

DR WPI; 1986-094065/14

[illegible]

PT	poly:peptide(s)	having antigenicity of RESA or FIRA antigens of
PT	falciparum	

PS Disclosure; Fig 7; 55pp; English.

CC The inventors claim a novel DNA molecule which comprises a
CC nucleotide sequence corresp. to all or a portion of the base
CC sequence coding RESA (AA060472) or FIRA (AA060473). RESA and FIRA have
CC antigenicity suitable for providing protective immunity against
CC *Plasmodium falciparum* malarial infections.

SQ Sequence 844 AA;

Query match 8.48; Score 553.5; DB 7; Length 844;

Matches 249; Conservative 117; Mismatches 359; Indels 209; Gaps 49

```

QY      19  IQQVSSVADNDKNK--RTKKKPRPKPVYDEAGSGLDNDFKVTYTPDTSTGHNNKVSISPK 76
      : : | | : : | : : : | : : : | : : | : : : | : : | : : : | : : | : : : |
Db      14  vdeyslsismenpqtatltlnlprdglsalnahlpn-einleltstlttnenevnpI 72
QY      77  TTTAPRLNRPISLPNS-----DYSKETSLVKNKETTVEKTTTNKQTSIDQKE 127

```

Db 638 paappappvratp-----ppappappn 665
 QY 839 SKATTPKPKP-----TKPKPTSTKKPTMPRVKPKTTTPPKMSTM-----PELNP 889
 Db 666 smalpappppppllaepappppplmpspappplppaapdpappplltlmgpspplap 725
 QY 890 TSRIEAMLQTTTRPNQTPNSKLVEVNPESDAGAGEGT 929
 Db 726 vpgaplapiingrpfvarknsl-----gsssgdt 756

RESULT 13

AAV53666
 ID AAV53666 standard; Protein; 4412 AA.

AC AAV53666;

DT 22-FEB-2000 (first entry)

DE Sequence g1/1017427/emb/CAA62189 from an alignment with protein 608.

KW Mechanical stress; gene therapy; protein 608; osteoporosis; bone density;

OS bone development; g1/1017427/emb/CAA62189.

XX Unidentified.

PN MO960164-A1.

PD 25-NOV-1999.

PE 14-MAY-1999; 99WO-US11066.

PR 15-MAY-1998; 98US-0085673.

PA (QUAR-) QUARK BIOTECH INC.

PI Elnat P, Mor O, Skalliter R, Feinstein E, Faerman A;

DR WPI: 2000-053304/04.

PT Identification of stress induced genes for determining risk and

PT preventing, treating or controlling osteoporosis

XX Claim 32; Fig 6A-R; 308pp; English.

CC The present sequence is obtained from a clustral X alignment with
 CC protein 608. Protein 608 was identified using the method of the invention
 CC after subjecting rat osteoblasts to mechanical stress. Expression of the
 CC 608 gene was found to be upregulated by about 3-fold in cells subjected
 CC to mechanical strain. The specification describes a method for the
 CC identification of genes responsive to a specific mechanical stress. The
 CC method comprises applying the mechanical stress to an organism (tissue
 CC or cells comprising bone cells), isolating the specific cellular
 CC fractions and extracting RNA from them, and differentially analysing the
 CC mRNA in comparison with control samples. The method is used to identify
 CC genes whose expression is responsive to a specific stress. The identified
 CC genes are employed in determining risk associated with a physiological or
 CC disease state. The risk determination methods are used for testing a
 CC medication for gene therapy. These medications, or genes identified by
 CC the method of the invention, are used for treating, preventing or
 CC controlling a physiological or disease state (especially osteoporosis or
 CC bone density or other factors causing or contributing to osteoporosis or
 CC its symptoms or other conditions involved in mechanical stress or its
 CC lack. The methods can also be used for advancing research or studies in
 CC bone development.

SO Sequence 4412 AA;

Query Match 9.5%; Score 625.5; DB 21; Length 4412;
 Best Local Similarity 26.2%; Pred. No. 8,3e-29;
 Matches 225; Conservative 86; Mismatches 379; Indels 169; Gaps 40;

QY 123 TDGKERTSAKENQSIK-----SAKDLPITSVLAKPTPKAETTTKGPALN----- 170
 Db 3515 tdfirgllqrveehvkvhrvleefeeaevefekkapp-----kpselsekllp 3566
 QY 171 TPKEPTPTTP-KEPATTPKEEPTPTTIKSAPTTTPEP-APTITKSAPTTKEPATTTTKE 228
 Db 3567 pkkpplkvvrkpeppkypevpkklvveekvrvpeepvppkypvevlppkvep----- 3621
 QY 229 PAPTTPKEPATTTTKEPATTTTKSAPTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKE 284
 Db 3622 ----ekkvvpvppkkrpeap-----ppkveapkvvevpkkyvpvppkkyvpvppkkyvpv 3672
 QY 285 TPPTTKEPATTTTKEPATTTT-----KEPATTPKAPPTTTPKEPATTTTKEPATTTTKE 337
 Db 3673 kaavpekkyvpealppkpspppevefeepespsaprk-kpevpvrvvpevpkvevpk 3731
 QY 338 EPSPTTTPKEPATTTTKSAPTTTKEPATTTTKSAPTTTKEPATTTTKEPATTTTKEPATTTTKE 397
 Db 3732 paap--pkkpevlrvk-vpeapkvvevpkkyvpv--ppkkyvpvptk-----vpevpvav 3782
 QY 398 PKRPAPPTTPKEPATTTTKEPATTTTTPKEPATTTTTPKEPATTTTTPKEPATTTTTPKEPATTTT 453
 Db 3783 pekkyv-----ealppkpspppevefeepespsaprk-kpevpvrvvpevpkvevpk 3838
 QY 454 PEKLAFTT-----PEKAPTTPEELAPTTPEEPTTP-TPPEEPATTPPKAAPNTPKEPAP- 507
 Db 3839 pekkapavvakkelppvkvvevpkvevpkkyvpv--peappkkyvpvkevpv 3896
 QY 508 ---TPKEP-----APTTPKEPATTTTKEPATTTTTPKEPATTTTTPKEPATTTTTPKEPATTTT 558
 Db 3897 kkvavpkkpevpkkyvpvkv-----leekpavp---vpeaespppevee--peel 3947
 QY 559 APTTTPKEPATTTTSDKAPTTTTPKEPATTTTTPKEPATTTTTPKEPATTTTTPKEPATTTTTPKEPATTTT 618
 Db 3948 ap---eeelapeekxpyvae-----eepevpvppavpeepkklipkkyvp-vlkkpeapp 3999
 QY 619 PKRPAPPELAPTTTTPKEPATTTTSDKAPTTTTPKEPATTTTTPKEPATTTTTPKEPATTTTTPKEPATTTT 669
 Db 4000 pkepepek-----lekplkprpppppppkedvkekllqkaidkkyvpvppkkyvpv 4056
 QY 670 TPTEPTPTTSEVS--TPPTTKEP-----TTIKSPDESPTPELSAETPALENSKE 720
 Db 4057 lrpikvpggekkvrrkliprpkpekeevlkkvllkrpeepkve---pkkle-kvkk 4111
 QY 721 PGVPTTKTPAAKRP-----EMTTADKTTTENDLRTPTTTPAAAPKMTKETATTTETTES 776
 Db 4112 pavp---eppppkpvveevpvlvkrerkllpeklvpeklpalipapepdkpkea----- 4164
 QY 777 KITATTTQVTSSTTQDTPPFKITTTTTLTKTTTLAPKVTTKTTTTEIMNK-----PESTA 830
 Db 4165 -----evklkppvpepplpaavvcpvuvvkaakapkeaeaa 4204
 QY 831 KPDATNSKATTPKPKPTAKPKKPTS--TKKRTMPRVKRPPTTTP-----RKMTST 883
 Db 4205 kpxgikvypktpspileaeerklrpsggekppdeapflqllkavplkfvkelkdlilt 4264
 QY 884 MELNPTSRIAEAMLQTTT 902
 Db 4265 esefvgssafecivpspt 4283

RESULT 14

AAW31855
 ID AAW31855 standard; Protein; 572 AA.

AC AAW31855;

DT 27-APR-1998 (first entry)

DE Mycobacterium tuberculosis 55 kDa protein.

XX Tuberculosis; mycobacteria; infection; diagnosis;

```

QY 497 AAPTNT---PKREPAPTTPKREPAPTTPKETAAPTTKGAAPTTLKEPAP----- 547
DB 640 anektssaeptengqrfanektssaeptehgerfplanenttlspaepnter 699
QY 548 TTPKPAPELAPTTKEPTSTTSKPAFTTTPKGAFTTTPKEAPTTKEPAPT-----T 602
DB 700 tanektfpaeptenreranentfspaqlengdrtplanektlsaeptengkrt 759
QY 603 PKGTAFTTLKEPAPT-----TPKKPA-PKE-----LAPTTTGP 635
DB 760 pfanektssaeptehaerfplanentsspaepnterentanektqfpaepntes 819
QY 636 TSTTSKPAFTTPEKT-----APTTKEPAPTTPKKPAFTTTPPTSEVST 684
DB 820 tanektfpaeptenrewtanenttlspaepthheemfplanektlsaeptengert 879
QY 685 PTTTKEPTTTHKSPDE-----STPELSAEPPTPK-----ALENSPEKPEQVPT---- 725
DB 880 pftnektlpsaeptehgerfplanektips-raeptehgerianekaklpsaeptehge 938
QY 726 -----TKTPAATKP-----EMTTAKDKTTERDLRTTPEPTTAPAKMTKE----- 765
DB 939 tlvedctpsaeptengertfplanenttlspkeshgerfentanektlpsaeptehger 998
QY 766 TATTTKTESKITAAT-TQVSTTTQDTTPFKI-----TTL----- 801
DB 999 tpsaneektlpsakptheemfpanenttlpsvpykthegektllanektlspepgleh 1058
QY 802 --KTTTLAPKYTTT-----KRTTTEIIMNKPEETAKPKDRAINSKATTPKPKP- 849
DB 1059 gaktlsaneektlpsakpthegeertspndktsaaestehdratsanvltipapepl 1118
QY 850 -----TKAPKKPSTKPKPTMPVRKP-----KTTPTP 877
DB 1119 khaktllahemktvteksstehekstste-ktttrekepllysektlctkgknprvp 1177
QY 878 RKMTSTPELNPSTSRIAEAMLTQTTTPNOT 907
DB 1178 ekptenlgnltlttelktkapvksstenpekt 1207

RESULT 12
AAW31852
ID AAW31852 standard; Protein; 763 AA.
XX
AC AAW31852;
XX
DT 27-APR-1998 (first entry)
XX
DE Mycobacterium tuberculosis 74 kDa protein.
XX
KW Tuberculosis; mycobacteria; infection; diagnosis;
KM antimycobacterial; antibiotic; vaccine.
XX
OS Mycobacterium tuberculosis.
XX
PN MO9741252-AA2.
XX
PD 06-NOV-1997.
XX
PE 18-APR-1997; 97MO-EP01973.
XX
PR 29-APR-1996; 96DE-4017184.
XX
PA (GBFB ) GBF GES BIOTECH FORSCHUNG GMBH.
XX
PI Espitia C, Honisch C, Moreno C, Singh M.
XX
DR WPI: 1997-549750/50.
XX
PT N-PSDB: AAT93610.
XX
PT New DNA and related proteins or RNA derived from M. tuberculosis -
used for diagnosis of mycobacterial infections, monitoring

```

```

PT vaccination and development of anti-mycobacterial agents
XX
XX Claim 5; Fig 13; 55pp; English.
CC This novel 74 kDa protein is encoded by an open reading frame of
CC a Mycobacterium tuberculosis DNA fragment (see AAT93610) containing
CC polymorphic GC-rich sequences. Its amino acid sequence shows
CC a high proline content, but there is no homology to any known
CC proline-rich antigens of mycobacteria. Novel M. tuberculosis
CC proteins (see AAW31851-57) are claimed. These can be produced as
CC recombinant proteins, especially in bacterial, yeast, fungal or
CC higher eukaryote host cells, and used for diagnosing tuberculosis
CC and other mycobacterial infections in humans or animals. The
CC claimed proteins can also be used for epidemiological studies, for
CC monitoring vaccination, and for the development of vaccines and
CC anti-mycobacterial drugs.
XX
XX Sequence 763 AA:
SQ
Query Match 10.9%; Score 715; DB 18; Length 763;
Best Local Similarity 28.2%; Pred. No. 5.3e-35;
Matches 248; Conservative 39; Mismatches 361; Indels 232; Gaps 39;
QY 156 PPKAETTTKGPALTTPKEPTTTPKEPASTTPKEPTTTSKAPTTPKREPAPTTKSAP 215
DB 3 PVP-----apralaplpappapaeakskpfiprap-----papcmmlvasap 46
QY 216 TTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKSAPTTPKREPAPTTKPAFTTPE 275
DB 47 pcp--pap-----papkpkakapifvpppaparelaplp--pap-----peagre 90
QY 276 PAPTTKEPTTPTT---PKREPAPTTKEPAPTTK-EPAPTPAK-KRAPTTKEPAPTTPK 329
DB 91 srpalpccpppvpvlpdppeapapvpnapspflfpfpaklvpapvp--pvpnspp 148
QY 330 EP--APTTKEPPTTPKREPAPTTTKSAPTTKEPAPTTKSAPTTPKESPPTTKEPAP 387
DB 149 ffpfpaaalpnap-----paplpanspplpappcpact--ppaapwpvpaakskpa 201
QY 388 TTPKEPAPTTPKKPAFTTTPKEPAPTTTKEPAP-----TTTKKAPNAPKRAPATTTPKENAP 443
DB 202 spptppap-----pmbatpmpfpplpvpdpdketlpapapaplpipppavipvpplp 256
QY 444 TTPKLTPTTPEKIAF-----TTPKEPAPTTPEELAPTTPEEPTPTTPEPAPTT-- 493
DB 257 pvpaklppapp--appvavaavlvapcpplpplpmnhppapapavpvgvlpaplpsnsh 313
QY 494 ---PKAAPNTPKEPAP----- 507
DB 314 papsapvpvgvlpaplplsgprvsvwkgstlslfocivcsgevlagalnsrpsrpl 373
QY 508 --TTPKEPAPTTPKREPAPTTPKETA-----PTTLKEPAPTTT-PKPAPE 557
DB 374 ttppalpaplpplpplpplpincavpplpplpvtalappplpplpplpplpplpplp 431
QY 558 LAPTTKEPTSTTSKPAFTTTPKGTAPTTPKERA-----PTTPKEPAPTTKGTAFTTL 611
DB 432 -pippgkpwcttpplapappppk-tvpvlppgpcpcpseknpdpapppeppksspalp 488
QY 612 KEAPPTT---KKPAPELAPTTTGP-TSTTSKPAFTTPEKT-----APTTKEPAPTTP 664
DB 489 appapmpsaavvppspplppapapaaasmpalppapspapalrcplppspapasp 548
QY 665 KKPAPTTTPEPTTSEVSTPTTKEPTTIHKS-----DESTPELSAEPPTKA--LENSP 718
DB 549 --pappapclpklis--anpccpvpnapnrtppapppapppelapapdpclpvpvansp 604
QY 719 KEQVPTTTPATKPEMTTAKDKTTERDLRTTETTTAARPKMKKEATTTTEKTESKI 778
DB 605 --pappapppapsalpfvnpa-----pplpaapk-----srpal 637
QY 779 TATTOVSTTTQDTTTPFKITTLKLTTLAPKVTTTKKLTITTEIIMNKPEETAKPKDRAFN 838

```


Best Local Similarity 38.5%; Pred. No. 6.7e-51;
Matches 226; Conservative 24; Mismatches 293; Indels 44; Gaps 11;

[illegible]

RESULT	10
AAW43106	
ID	AAW43106 standard; Protein; 1664 AA.
XX	
XX	AAW43106;
AC	
XX	16-OCT-1998 (first entry)
DT	
XX	
DE	C. thermocellum O1pb protein.
XX	
KM	Multimer; enzyme; complex; protein-protein interaction; dockerin domain;
KM	cohesin domain; catalytic subunit; scaffold subunit; SdBA; synergistic;
KM	cellulosome integrating protein; scaffoldin dockerin binding protein.
XX	
OS	Clostridium thermocellum.
XX	
FH	
Key	Location/Qualifiers
FT	28..192
FT	/note= "cohesin type II domain"
FT	207..363
FT	/note= "cohesin type II domain"
FT	409..565
FT	/note= "cohesin type II domain"
FT	607..763
FT	/note= "cohesin type II domain"
XX	
PN	FR2748479-A1.
XX	
PD	14-NOV-1997.

XX 10-MAY-1996; 96FR-0005854.
PF
XX
PR 10-MAY-1996; 96FR-0005854.
XX
PA (INSP) INST PASTEUR.
XX
PI Beguin P, Leibovitz E;
XX
XX
XX
DR WPI: 1998-01569/02.
DR N-PSDB: AAT86623.
XX
XX
XX Cellulase proteins with cohesin or dockerin type II domains - useful
PT for potentiating the activity of multiprotein enzyme complexes
XX
XX Claim 7, Page 31-39; 60pp; French.

XX Multimeric protein, especially enzymatic, complexes are held together
CC by protein-protein interactions between domains designated dockerins
CC and cohesins, which are found on the catalytic and scaffold subunits
CC respectively. An example of such a complex is the cellulose degrading
CC protein complex from *Clostridium thermocellum*, known as the cellulosome.
CC This complex comprises around 15 proteins including endoglucanases,
CC cellobiohydrolases, hemicellulases, e.g. xylnases or lichenases, which
CC interact with a central "scaffold" protein designated the cellulosome
CC interacting protein (CIP; see A0443108). The catalytic subunits
CC interact with the CIP subunit via conserved 23 amino acid dockerin
CC domains. CIP has been shown to contain 9 copies of a cohesin domain.
CC The invention relates to the isolation of proteins binding to a novel
CC dockerin type domain found in the C-terminal portion of CIP. The new
CC domain is designated a type II dockerin domain (as compared to the type
CC I domain found on the catalytic subunits of the cellulosome). The
CC II dockerin domain has some sequence similarity to the type I dockerins
CC but is unable to bind type I cohesin domains.
CC The sequence presented here is an example of a protein which binds
CC the novel type II dockerin domain and is the product of the OIPB gene.
CC The protein contains 4 type II cohesin domains in the N-terminal portion
CC of which the first domain (amino acid residues 28-192) is thought to
CC bind CIPa. The novel type II dockerin and cohesin domains can be used
CC in complexes, especially enzyme complexes, to potentiate their catalytic
CC actions in a synergistic manner.

[illegible]

PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	21-JUN-1999;	99US-0139752.
PR	22-JUN-1999;	99US-0139817.
PR	23-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	24-JUN-1999;	99US-0140354.
PR	28-JUN-1999;	99US-0140695.
PR	29-JUN-1999;	99US-0140823.
PR	30-JUN-1999;	99US-0140991.
PR	01-JUL-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	02-JUL-1999;	99US-0142154.
PR	06-JUL-1999;	99US-0142055.
PR	08-JUL-1999;	99US-0142390.
PR	09-JUL-1999;	99US-0142803.
PR	12-JUL-1999;	99US-0142920.
PR	13-JUL-1999;	99US-0142977.
PR	14-JUL-1999;	99US-0143542.
PR	15-JUL-1999;	99US-0143624.
PR	16-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	21-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	26-JUL-1999;	99US-0145224.
PR	27-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	03-AUG-1999;	99US-0146389.
PR	04-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148319.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148604.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149375.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	23-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	25-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	27-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	30-AUG-1999;	99US-0151080.
PR	31-AUG-1999;	99US-0151303.
PR	01-SEP-1999;	99US-0151338.
PR	07-SEP-1999;	99US-0151930.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154079.
PR	20-SEP-1999;	99US-0154773.
PR	22-SEP-1999;	99US-0155118.
PR	23-SEP-1999;	99US-0155138.
PR	24-SEP-1999;	99US-0155486.
PR	28-SEP-1999;	99US-0155659.
PR	29-SEP-1999;	99US-0156458.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	18-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	22-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	25-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	26-OCT-1999;	99US-0161405.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match

14.9%; Score 981; DB 21; Length 763;

D0 2128 t---tptttcttcprbpccgqccpcrccccccc-ct-t-t-g-a-a-

243 dymrrrrdD---tttb--tTTTAAAPKMT---KETATTEKTESKITATTQVSTTTQDP 7

PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.

27-MAY-1999; 99US-013678
28-MAY-1999; 99US-013679
01-JUN-1999; 99US-013722


```

XX OS Homo sapiens.
XX FH Key
XX FT Misc-difference 393.396
XX FT /note= "unspecified amino acids"
XX FT Misc-difference 444.446
XX FT /note= "unspecified amino acids"
XX PM W09523861-A1.
XX PD 08-SEP-1995.
XX PE 06-MAR-1995; 95MO-CN00015.
XX PR 04-MAR-1994; 94CN-0112066.
XX PA (SHAN-) SHANGHAI BEITI BIOTECHNOLOGY CO LTD.
XX PI Gu X, Han Z, Shen Q.
XX DR WPI: 1995-320576/41.
XX DR N-PSDB: AAT04546.
XX PT New haematopoietic cell growth factor - used for treating
XX PS thrombocytopenia and hematocytopenia
XX PS Example: Page 23: 36pp; Chinese.
XX CC This sequence represents the human megakaryocytopoietin (MPO) protein.
XX CC This sequence was purified using a carrier which can couple wheat germ
XX CC agglutinin and heparin to separate MPO. Fragments of this sequence (see
XX CC AAR80039 and AAR80040) were used to produce the amplification primers
XX CC shown in AAT04544 and AAT04545. The fragments amplified by these primers
XX CC can then be used as probes to screen human cDNA libraries for MPO cDNA.
XX CC The MPO cDNA can then be inserted into a plasmid which is used to
XX CC transform cells to produce MPO. The MPO sequence is capable of promoting
XX CC colony formation of megakaryocytes, enlarging the size of megakaryocytes
XX CC and stimulating the proliferation of multipotential stem cells. The
XX CC factor may be used for treating thrombocytopenia and hematocytopenia.
XX CC The purification method can be used to isolate MPO from human urine or
XX CC serums of patients with aplastic anaemia, and from animal blood or urine
XX CC by radiation exposing the animals to induce aplastic anaemia.
XX SO Sequence 452 AA;

```

```

Query Match 21.3%; Score 1397.5; DB 16; Length 452;
Best Local Similarity 63.2%; Pred. No. 4.8e-76;
Matches 287; Conservative 3; Mismatches 21; Indels 143; Gaps 2;

```

```

QY 1 MARRTLPIYLLLSVIVIOVSS----- 24
DB 1 MARRTLPIYLLLSVIVIOVSS----- 24
QY 25 ----- 24
DB 61 kvctaelsockgrcfesfgrccodagckkydcocpdyesfcaevnmpspsskkap 120
QY 25 ----- 24
DB 121 ppsgasqtlkctckrpyhpnkkkkkyseellevdnknrtkkkprkprvdeag 48
QY 49 SGLDNGDFVYTPDSTTOHNKSVSPKITTAKPINPPSLPPNSDSKESSTLVNKEET 108
DB 181 sgldngdfkvttptdsttqnhkvstpslttackpnpnpslppnsdskselstlvnket 240
QY 109 VETKETTNNKOTSDGKEXKTTSAKETOSIKETSAKDLAPTSVLAAPTPKAETTTGPA 168
DB 241 vetkettnnkvstcdgkextsketstaketsakdlaptsvklakptpkkaetttgpa 300
QY 169 LTTPEKPTPTPKPEASTPTPKPTTTSKAPTTPKPEAPTTSATPTTKEPAPTTKE 228
DB 169 LTTPEKPTPTPKPEASTPTPKPTTTSKAPTTPKPEAPTTSATPTTKEPAPTTKE 228

```

```

DB 301 lttpeptptpkpeasttpkpeptltksaptckpeaptltksaptltkpeapttke 360
QY 229 PAPPTPKPEAPTPTTKEPAPTPTTTSAPPTPKPEAPTPTTTPKAPPTTKEPTPTT 288
DB 361 papptpkpeaptltkpeaptltkshpplprscxxxxctgp-----tpkshppl 409
QY 289 PREPAPTTPKEPAPTTPKPEAPTAPKAPPTTPE 322
DB 410 prshpckpeaptltkpeaptltkshpplprsc 443

RESULT 8
AAM24516
ID AAM24516 standard; protein: 5179 AA.
XX AC AAM24516;
XX AC 12-OCT-2001 (first entry)
XX DE C899P predicted amino acid sequence.
XX KM Human: immunotherapy; diagnosis: colon cancer; colon tumour;
XX KM immunogenic; gene therapy; vaccine; colonic cancer.
XX OS Homo sapiens.
XX PA Homo sapiens.
XX PM W0200149716-A2.
XX PD 12-JUL-2001.
XX PF 29-DEC-2000; 2000MO-US35596.
XX PR 30-DEC-1999; 99US-0476296.
XX PR 10-JAN-2000; 2000US-0480321.
XX PR 13-FEB-2000; 2000US-0504629.
XX PR 06-MAR-2000; 2000US-0519444.
XX PR 19-MAY-2000; 2000US-0575251.
XX PR 29-JUN-2000; 2000US-0609448.
XX PR 28-AUG-2000; 2000US-0649811.
XX PA (CORI-) CORIYA CORP.
XX PI Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
XX PI King GE, Wang T, Jiang Y;
XX PF WPI: 2001-441847/47.
XX PS Colon tumor associated proteins and nucleic acids useful for the
XX PS prevention, diagnosis and treatment of colonic cancer -
XX PS Claim 2; Page 446-462; 472pp; English.

The present invention describes colon tumour associated proteins (I) and
the polynucleotides (II) that encode them: (I) have cytostatic activity.
(I) and (II) can be used in gene therapy and vaccine production. (I) and
(II) may be used in the prevention, diagnosis and treatment of diseases
associated with inappropriate colon tumour associated protein (TCAP)
expression, such as colonic cancer. For example, (I) and (II) may be
used to treat disorders associated with decreased expression by
rectifying mutations or deletions in a patient's genome that affect the
patients own production of them. Additionally, (II) may be used to
produce the TCAP proteins, by inserting the nucleic acids into a host
cell culturing the cell to express the protein. (II) and its
complementary sequences may also be used as DNA probes in diagnostic
polymerase chain reaction (PCR) and hybridisation assays to detect and
quantitate the presence of similar nucleic acids in samples, and
therefore which patients may be in need of restorative therapy. (I) may
also be used as antigens in the production of antibodies against TCAPs
and in assays to identify modulators of TCAP expression and activity.
Anti-(I) antibodies and antagonists may also be used to down regulate
TCAP expression and activity. The anti-(I) antibodies may also be used
as diagnostic agents for detecting the presence of TCAPs in samples

```

PT	New composition comprising the camptodactyly-arthropathy-coxa
PT	vara-pericarditis protein in combination with an anesthetic, useful for
PT	treating osteoarthritis, or as lubricants of tissue and joints -
XX	Example 1; Fig 4; 34pp; English.
XX	The invention relates to a method of treating osteoarthritis via the
CC	administration of a composition comprising the camptodactyly-arthropathy-
CC	coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.
CC	The composition may further comprise a local anesthetic. The composition
CC	of the invention may be administered via intra-articular or intravenous
CC	injection. The human CACP protein is identified in the invention as
CC	being megakaryocyte stimulating factor (MSF). The gene encoding in
CC	CACP protein (MSF) is located on chromosome 1q25-31, and mutations in
CC	this gene are responsible for the heritable disorder camptodactyly-
CC	arthropathy-coxa vara-pericarditis, in which patients have synovial
CC	hyperplasia without evidence of inflammation. CACP protein (MSF)
CC	acts as a synovium lubricant, and can be used to lubricate tissue and
CC	joints in the treatment of osteoarthritis. The composition may be
CC	applied to reduce the symptoms of osteoarthritis (e.g., joint pain,
CC	loss of range of movement or joint damage). The present sequence
CC	represents the bovine orthologue of human MSF, superficial zone
CC	protein (SZP).
CC	
XX	Sequence 472 AA:
SQ	
OY	832 PKDRATNSKATTPPKOKPTAPKKPPSTKKMKMPVRPRKPTTPTPRRK-MTSMPELNPT 890
Db	76 pkgratsqyltpkpkdkpkapkprtsckkprt-prvkpkttptpptktsampercpt 134
OY	891 SRIDAMIQTTTRNQPNPSKSLVEVNRKSDAGAGEPRHMLLRPHVMPEYTPMDL 950
Db	135 s-lpeamqtcttptbpnseilidnsenedgdaagekhmlfipvlpbiivgtell 193
OY	951 PRVNOGIINPMUSDETNICNGKPVOGLTIIRNGTLVARGHVFWMLSPPSPSPARRI 1010
Db	194 vrgpqdgfgjpmfdesdenlcngrpvddglitllngallvaifghyfmmultftfpppprrl 253
OY	1011 TEWNIIPSPDITVPFRNCCEKTEFFFKDSQRYWRFTNDIKDGYPKPIFGGIGTGIVA 1070
Db	254 tewwvipdvtvftrcnccegfktffkgsgywrffndikdagypkliskgfgingtkfa 313
OY	1071 ALSTAKYNMPESVYFEKRGSGTOQYITKOEVYQCKCPBRBALNVPYGVGMTQVRRRFE 1130
Db	314 alslaqkykerpsvyffkigsgsvgylykdeqltcvgripingysygetlaqrtrrffe 373
OY	1131 RAIGSQHTTIRIOYSPARLAYODKGVLHNKEVYSILMRGLPNVTSAISLPNIRKPNYG 1190
Db	374 ralspgqhtltirhycpvrpygdgflhnekvkstlwrqlpnvvtsaislpnlrkpcgy 433
OY	1191 DYAFESKDQYNNIDVPSRTARAATTTRSQTLSKWNYNCP 1229
Db	434 dyafelskdqymnidvpstaraatttrsqtlsnwynpc 472
RESULT 7	
AAR80041	ID AAR80041 standard; Protein; 452 AA.
AC	AAR80041:
DT	10-APR-1996 (first entry)
XX	Human megakaryocytopoietin protein.
DE	
XX	Human; megakaryocytopoietin; wheat germ agglutinin; heparin;
KM	megakaryocyte; aplastic anaemia; thrombocytopenia; hematocytopenia;
KM	multipotential stem cell.

Db	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465
----	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

PR	03-AUG-2000; 20000US-0631451.
PR	15-SEP-2000; 20000US-0663870.
XX	
XX	(HXSE-) HXSEQ INC.
PA	
XX	
PI	Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI	Cao Y, Drmanac RA, Zhang J, Werhman T;
XX	
DR	WPI; 2001-476164/51.
XX	N-PSDB; AAH98981.
XX	
PT	Isolated polypeptide for treatment of diseases, diagnostics, raising
PT	antibodies and research use -
XX	
PS	Claim 20; Page 1198-1201; 1275pp; English.
XX	
CC	The present invention provides the protein and coding sequences of novel
CC	proteins from a variety of organisms, including human, dog, cat, horse,
CC	cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC	urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC	from the organism of interest. They can be used in diagnostics,
CC	forensics, gene mapping, identification of mutations, to assess
CC	biodiversity and for nutritional purposes. The present sequence is a
CC	protein of the invention.
XX	
XX	Sequence 1299 AA;

	Query Match	89.7%;	Score 5894.5;	DB 22;	Length 1299;	
	Best Local Similarity	86.5%;	Pred. No. 0;	Mismatches	0;	Indels 175; Gaps 1
	Matches 1120;	Conservative	0;			
OY	1 MAWKTEPIYLLLLLSEFVIOOVSS-----					24
Dd						
	1 mawktlpiy11111svrviqavssgdlsscagcgegyrdatcncdyngchymccpdf					60
OY	25 -----					24
Dd	61 kvvctaelsckycrfesfergrecedaqckyydkccpdyesfcaeavnmp-spskskap					120
OY	25 -----					24
Dd	121 ppsgasqlkctkrppkpnnkkkkkvieslethehsvenqeasssssssstlw					180
OY	25 -----VKDKNNKTKKKPKPPPYDEAGSGDNGDFVTVPDTST					65
Dd	181 k1skksnaanrelqklkvndknkntkkkbpkpvvdeagsjldngdfvtvpdst					240
OY	66 TQHNKVYSNPKTTAKPINRPSLPBNSDSKETSLTNKETTVEETKNTTNNKOTSDG					125
Dd	241 tqnkvysnpklttckpnlprpslpnsdstsketslvnkettveketllnkqstdg					300
OY	126 KEKTTSAKETQSIEKTSAKDLAPISKVLAKTPAEKTTKGPAITPKEPTTPKEPAS					185
Dd	301 kekttaketqsiekttsakdlapiskvlyaktpaeelttgpalitpkcpttkepas					360
OY	186 TPPEKPTPTTISAAPTKEPAATTTKSAPPTKEPAATTTKEAPAATTTKEPATTTKBP					245
Dd	361 ttpkeppttiksappttkcpapttkcappttkcappttkcappttkcappttkcap					420
OY	246 APPTTASAPTTKEPAATTPKKRAPTTTTKEAAPTTPKPPTTTPKEBAPTTTKEBAPTTPK					305
Dd	421 apttkasapttkcppttkcpkppapctpkcppttkcppttkcpapttkcappttkp					480
OY	306 EPAPTAKKRAPPTPEKAAPTTPKEBAPTTTKEPSPTTKEBAPTTTKSATTTKEBAPT					365
Dd	481 epaptakkrappttkcpapttkcpapttkcpspttkcpapttkcsapttkcappt					540
OY	366 TTGSAPTPKEPSPTTKEPAATTPKEBAPTTPKKAPATTPKEBAPTTPKEBAPTTTKKP					425
Dd	541 ttgsaptpkpspttkcpapttkcpapttkcpapctpkpkaapttkcpapttkcpapttkkp					600
OY	426 APLTAPEAPLTPPEAPPTPKKLTPTEKTLAPTTPEKPADTTPBELAPTTPEEPPTTT					485

```
OY 846 POKRTAKPKPTSTKPKPTMPRVKRRKPTTPTPKMTSTMBELNPTSRIAEAMLOTTBRN 905
    |||||||
Db 1021 pqrpkckpkpkstckkpkckmpvrkdktptrckmtstmpelmptriseamllqcttprn 1080
OY 906 QTNNSKLVENPKSEDAAGGEGETPMMLRPHVFMPEVTPDMDYLRVNOGIIINPMLS 965
    |||||||
Db 1081 qtnpsklvenpksedagaggetpmlilrphvfmpvdpmdyldprvpngqilinpmls 1140
OY 966 DETNINCKRPVDGLTTLRNGTLEAFNGHYFMMLSPSPSPARITTEWGCISPIDYVPT 1025
    |||||||
Db 1141 detnlcngrpvdglttlrnlgtlvaftgghyfmmlspspsparrtlewglpsldvft 1200
OY 1026 RCNCEKTEFFRDSQYWRRTNDIKAGYKPIFKGFGGLTGQIYVALSTAKYKNPESY 1085
    |||||||
Db 1201 rcncgkctffkdsqgwrftndikdagypkpfkfgfgltgqivaalstakympeasy 1260
OY 1086 FFRKGSIOOYIYKQEPVQKCPGRPALNYPYGEMTOVRRRRERRAIGPSTHTIRIQY 1145
    |||||||
Db 1261 ffrkgsiqgylqkqepvqkcpgrrrpalnypvgemtvrrrrferralgpsqthlrriqy 1320
OY 1146 SPARLAYQKGVILHNEVKVSIIMRGLPNVVTSAISLPNIRKPDGYDYAFSKDQYNNIDY 1205
    |||||||
Db 1321 sparlqyqkgyilhnevkvslmrqlpnvvtalslpnirkrpdydyafskdqynldv 1380
OY 1206 PSRTARAITTRSGOTLSKWNWYCP 1229
    |||||||
Db 1381 psrtaraittrsqtlskwnwncp 1404
```

RESULT 3

AA029773
ID AA029773 standard; Protein: 1404 AA.

AA029773;

DT 28-FEB-2001 (first entry)

XX Human megakaryocyte stimulating factor (MSF), SEQ ID NO:1.

XX Human MSF, megakaryocyte stimulating factor; tribonectin;

KW alternative splicing; joint boundary lubricant; O-linked oligosaccharide;

KW osteoarthritis; tribosupplementation; tissue adhesion inhibition;

KW friction coefficient reduction; gene therapy; antiarthritic;

KW osteopapalic.

XX Homo sapiens.

OS MO200064930-A2.

PN 02-NOV-2000.

PD 24-APR-2000; 2000MO-US10953.

PF 23-APR-1999; 99US-0298970.

PR (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.

XX Jay GD;

XX WPI; 2001-024673/03.

DR N-PSDB; AAC81498.

XX Novel tribonectin polypeptide useful as lubricant for treating
PT osteoarthritis, comprises O-linked lubricating moiety
XX
PS Claim 3; Page 7; 47pp; English.

CC The invention relates to a human tribonectin which is a product of
CC alternative splicing of the human MSF (megakaryocyte stimulating factor)
CC gene. The tribonectin has at least one O-linked oligosaccharide
CC lubricating moiety and has a polypeptide sequence comprising 1-76
CC repeats of a motif having at least 50% identity to the sequence KRPAPPT
CC (AA029774). The invention also relates to a nucleic acid encoding a

```
CC human MSF-derived tribonectin; a biocompatible composition comprising a  
CC human tribonectin for inhibiting tissue adhesion formation; and a method  
CC of diagnosing osteoarthritis or a predisposition to osteoarthritis by  
CC measuring the amount of MSF or its fragment in a biological sample of a  
CC mammal, wherein an increased amount of MSF compared to a control  
CC indicates the presence of or predisposition to developing  
CC osteoarthritis. The tribonectin and DNA encoding it are useful in the  
CC treatment of osteoarthritis, where they may be used for lubricating  
CC mammalian joints, such as articulating joints of humans, dogs or horses.  
CC The tribonectin, when formulated as a membrane, foam, gel or fibre, is  
CC useful for inhibiting adhesion between two surfaces such as the injured  
CC or trauma, or an artificial device e.g., an orthopaedic implant. In  
CC particular, one of the surfaces is pericardial tissue. DNA encoding a  
CC human MSF.  
XX  
SQ Sequence 1404 AA:
```

Query Match 98.5%; Score 6470.5; DB 22; Length 1404;
Best Local Similarity 87.5%; Pred. No. 0;
Matches 1229; Conservative 0; Mismatches 0; Indels 175; Gaps 1;

OY 1 MAMKTLPIYELLLSVFVQVSS-----VKDNKKRTRKRRKPTPKPPVDEAGSLDNGDKRYTTPDST 24

Db 1 mawktlpiyllllsvfvlqvsqdlsscagrcgegyardatcnodyncghymecpof 60

OY 25 ----- 24

Db 61 kvctaelockgrcfesfergrecddagckkydkccpdyesfcaevhnptspaskkap 120

OY 25 ----- 24

Db 121 ppsgaqltkstckrpkpknkkkkykyleseelceehsvsengessssssssstlw 180

OY 25 -----VKDNKKRTRKRRKPTPKPPVDEAGSLDNGDKRYTTPDST 65

Db 181 klksksnaaneelkklklyknknkrrtkkpcbpvvdagsgldgdkvtvtpdst 240

OY 66 TQHNKVTSPKTTTKPPINRPSLPNSDTSKENSILYVNEKTYETKTTNNKOTSDG 125

Db 241 tqhkvstspkltaklpnrpslpnsdtskelslvnketvckettlnqtsldg 300

OY 126 KEKTSAKETOSIEKTSADLAPTSKVLAKPPRAETTTKGPALTTEKEPTTPKEPAS 185

Db 301 kektsaketgsiektksakdlaptskvlakpbraettkgpalttkepttkepas 360

OY 186 TTPKEPTPTTKSAPTPKEPAPTTKSAPTPKEPAPTTKSAPTPKEPAPTTKKEP 245

Db 361 ttpkeptpttksaptpkepafttksaptpkepafttksaptpkepafttkkepa 420

OY 246 APTTKSAPTPKEPAPTTKKRAPPTTKKRAPPTTKKRAPPTTKKRAPPTTKKRAP 305

Db 421 apttksaptpkepafttkkrapfttkkrapfttkkrapfttkkrapfttkkrap 480

OY 306 EPAPTPKRAPPTTKKRAPPTTKKRAPPTTKKRAPPTTKKRAPPTTKKRAPPTTK 365

Db 481 epaptpkrapfttkkrapfttkkrapfttkkrapfttkkrapfttkkrapfttk 540

OY 366 TTKSAPTPKEPSPTTKKRAPPTTKKRAPPTTKKRAPPTTKKRAPPTTKKRAP 425

Db 541 ttksaptpkepspttkkrapfttkkrapfttkkrapfttkkrapfttkkrapfttk 600

OY 426 APTAPKEPAPTTKKRAPPTTKKRAPPTTKKRAPPTTKKRAPPTTKKRAPPTTK 485

Db 601 aptapkepafttkkrapfttkkrapfttkkrapfttkkrapfttkkrapfttk 660

OY 486 PEAPPTTKKRAAPNTPKEPAPTTKKRAPPTTKKRAPPTTKKRAPPTTKKRAP 545

Db 661 peepaptpkrapnptkrapfttkkrapfttkkrapfttkkrapfttkkrapfttk 720

OY 546 APTTPKKRAPKELAPTTTKEPTSTSDKRAPPTTKOTAPPTTKRAPPTTKRAPPTTK 605

QY 1086 FFKKGGSIQOYITTKQEPVOKCPGRPALNYPVGEEMTOVRRRPERAIGPSQTHTRIQY 1145
 |||||||
 Db 1261 ftkrgsgyqylykqepqkcpgrpalnypvgeemqvttrrrtererlgsqtblrltlyq 1320
 |||||||
 QY 1146 SPARLAYODKGVILHNEVKSILMRGLPMWVTSALSPNIRKPDGYDYAASKOYINIDV 1205
 |||||||
 Db 1331 sparlaydkgvllhnevksllmrglpmwvtsalsplnirfxpdydyafakdgyndlv 1380
 |||||||
 QY 1206 PSRTARAITTRSGOTLSKWTNCP 1229
 |||||||
 Db 1381 psrtaraittrsgtllskwtyncp 1404
 |||||||

RESULT 2
 AAB60568
 ID AAB60568 standard; Protein: 1404 AA.
 XX
 AC AAB60568;
 XX
 DT 27-APR-2001 (first entry)
 XX
 DE Human megakaryocyte stimulating factor (MSF, CACP).
 XX
 KW Human: CACP protein; camplodactilyl-arthropathy-coxa vara-pericarditis;
 KW MSF: megakaryocyte stimulating factor; synovial lubricant;
 KW chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;
 KW anltrahitic.
 XX
 OS Homo sapiens.
 XX
 PN W0200107068-A1.
 XX
 PD 01-FEB-2001.
 XX
 PF 21-JUL-2000; 2000MO-US20002.
 XX
 PR 23-JUL-1999; 9905-0145328.
 PR 19-JUL-2000; 2000US-0145328.
 XX
 PA (UYCA-) UNIV CASE WESTERN RESERVE.
 XX
 Warman ML;
 PI
 XX
 DR WPI; 2001-182721/18.
 XX
 PT New composition comprising the camplodactilyl-arthropathy-coxa
 PT vara-pericarditis protein in combination with an anesthetic, useful for
 PT treating osteoarthritis, or as lubricants of tissue and joints
 PT
 XX
 PS Example 1; Page -: 34pp; English.
 XX
 CC The invention relates to a method of treating osteoarthritis via the
 CC administration of a composition comprising the camplodactilyl-arthropathy-
 CC coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.
 CC The composition may further comprise a local anesthetic. The composition
 CC of the invention may be administered via intra-articular or intravenous
 CC injection. The human CACP protein is identified in the invention as
 CC being megakaryocyte stimulating factor (MSF). The gene encoding
 CC CACP protein (MSF) is located on chromosome 1q25-31, and mutations in
 CC this gene are responsible for the heritable disorder camplodactilyl-
 CC arthropathy-coxa vara-pericarditis, in which patients have synovial
 CC hyperplasia without evidence of inflammation. CACP protein (MSF)
 CC acts as a synovium lubricant, and can be used to lubricate tissue and
 CC joints in the treatment of osteoarthritis (e.g., joint pain,
 CC applied to reduce the symptoms of osteoarthritis). The present sequence
 CC loss of range of movement or joint damage). The present sequence
 CC represents human megakaryocyte stimulating factor (MSF, CACP protein).
 CC Note: This sequence is not given in its entirety in figure 4 of the
 CC specification, although a Genbank accession number was given. This
 CC sequence was therefore obtained from Genbank (U70316).
 XX
 SQ Sequence 1404 AA;

Query Match 98.5%; Score 6470.5; DB 22; Length 1404;
 Best Local Similarity 87.5%; Pred No. 0;
 Matches 1229; Conservative 0; Mismatches 0; Indels 175; Gaps 1;

QY 1 MAKKTLPIYLLLSVFIQOVS----- 24
 |||||||
 Db 1 makktlpiyylillsvfivqgvsqdlssacgrcgysrdactncdyncqhywecpdlf 60
 |||||||

QY 25 ----- 24
 |||||||

Db 61 kvctaelseckgrcfesfergrecdcaqkkydkccpdyesfcaevhnlpspskcap 120
 |||||||

QY 25 ----- 24
 |||||||

Db 121 ppsgaagtlstkrspkpnkkkllkvieseeeltehsvensgeessssssstliw 180
 |||||||

QY 25 -----VKDNKKNRTKKKPPKPPVVDFASSGJDNDFKXTPTPTST 65
 |||||||

Db 181 kikesksaanrelqkklkvdknkrkllkppkppvveaagsqldnqdfkvlrpldst 240
 |||||||

QY 66 TQHNKVSTSEKITTAQINRPSLPNSDSKETSILVNKETTVEKETTNNKQSTSDG 125
 |||||||

Db 241 tqhnkvstspkiltakpimprslpnsdsksesltnketvetkettlnkqstsdg 300
 |||||||

QY 126 KEKTTSAKETOSIKETSAKDLAPTSTKVLAKPPKAETTTGPAITTPKEPTTPKEPAS 185
 |||||||

Db 301 kektsakeitgsiektsakdlaptskvlakpkaetltkypalittkptkptkpas 360
 |||||||

QY 186 TTPKEPTTPKSAPTTPKEPAPTTTSAPTTPKEPAPTTTPKEPAPTTTPKEP 245
 |||||||

Db 361 ttpkepttpktsaplttkaplttkaplttkaplttkaplttkaplttkaplttkap 420
 |||||||

QY 246 APTTKSAPTPPKPEAPTPPKKPAPTTPKEPAPTTPKEPTTPKEPAPTTTPKEP 305
 |||||||

Db 421 apttksaptpkpeaptpkpkpapttkpapttkpapttkpapttkpapttkpapttk 480
 |||||||

QY 306 EPAPTAPKPAPTTPKEPAPTTTPKEPATTTKESPTTPKEPATTTTSAPTTTPKEPAT 365
 |||||||

Db 481 epaptapkkpapttkpapttkpapttkpapttkpapttkpapttkpapttkpapttk 540
 |||||||

QY 366 TTKSAPTPPKESPTTPKEPATTPKEPATTPPKKAPATTPKEPATTPKEPATTTKKP 425
 |||||||

Db 541 ttksaptpkpespttkpapttkpapttkpapttkpapttkpapttkpapttkpapttk 600
 |||||||

QY 426 APAPTKEPATTPPKETAPTTPKKLTPTTPKEKLAPTTPKEPATTPPEELAPTTPPEEPTT 485
 |||||||

Db 601 apaptkepapttkpapttkpapttkpapttkpapttkpapttkpapttkpapttkpapttk 660
 |||||||

QY 486 PEEAPATTPKKAAPNTPKPEAPTPPKPEAPTPPKPEAPTPPKETAPTTPKGTAPTTPKEP 545
 |||||||

Db 661 peepaptpkkaapntpkpeaptpkpeaptpkpeaptpkpeaptpkpeaptpkgtapttk 720
 |||||||

QY 546 APPTPKKPAKELAPTTTKEPTSTSDKPAPTTPKGTAPTTPKEPATTPPKPATTTKPG 605
 |||||||

Db 721 aptpkpkpaelaptttkesptstsdkpaapttkgtapttkpapttkpapttkpapttkpg 780
 |||||||

QY 606 TAPTTLKEPATTPPKKPAKELAPTTTKEPTSTSDKPAPTTPPKETAPTTPKEPATTPPK 665
 |||||||

Db 781 tapttlkepapttkpkpaelaptttkesptstsdkpaapttkgtapttkpapttkpapttkpg 840
 |||||||

QY 666 KPAPTPEPTTPPTSEPTSTPTTKEPTTHKSPDESPELSAPPKALENSPKKEGVT 725
 |||||||

Db 841 kpaptpepttpptseptstpttkesptthkspdespeelsaepkalenspkkegvyt 900
 |||||||

QY 726 TKTPPATRPMPTTAAKDUTTERDLRTTPTTAAKMTKETATTTKTESKTTATTTTGV 785
 |||||||

Db 901 tktpaatrptmtttaaakduttrerdllrtptttaaakmtketatlttkteskttatltty 960
 |||||||

QY 786 TSTTQOTPTPKTTTLKTTTLAPKVTTTTKTTTTEIMNKPEETAPKPKDRATNSKATTPK 845
 |||||||

Db 961 tsttqotptpktttlktttlapkvttttktttiteimnkpeetapkpdatnskattpk 1020
 |||||||

FT	Region	1266..1331
FT	/label= Exon_X	
ET	Region	1331..1373
FT	/label= Exon_XI	
FT	Region	1373..1404
XX	/label= Exon_XII	
PX	WQ9213075-A.	
PN		
XX		
PD	06-AUG-1992.	
PF		
PE	17-JAN-1992;	92MO-USO0433.
XX		
PR	18-JAN-1991;	91US-0643502.
PR	10-SEP-1991;	91US-0757022.
XX		
PA	(GEMV) GENETICS INST INC.	
PI	Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K,	
PI	WPI, 1992-284660/34.	
DR	N-PsDB; AMQ27223.	
DR		
XX		
PT	New human mega-karyocyte stimulating factors - for treating	
PT	immune deficiencies, cancer, exposure to radiation or drugs,	
PT	bacterial and viral infections, etc.	
PS	Claim 1, 2 and 3; Fig 1; 87pp; English.	
XX		
CC	The sequence given is a full length translation from the megakaryocyte	
CC	stimulating factor (MSF) precursor. The sequence covered by exons II,	
CC	III and IV encodes megakaryocyte stimulating factor (MSF). This	
CC	sequence is modified by the addition of an N-terminal sequence encoding	
CC	a secretory leader, an initiating methionine preceding exon II and a	
CC	terminating codon following exon IV. The cDNA sequence given contains	
CC	sequences derived from human megakaryocyte colony stimulating factor	
CC	(meg-CSF). Exon I contains the initiating methionine, and encodes a	
CC	classical mammalian protein secretion signal sequence. The sequence	
CC	encoding the original meg-CSF includes exons II-IV and is thought to	
CC	terminate in the region between amino acid residues 114 - 147. The	
CC	primary transcript of this gene may be cleaved in different ways to	
CC	yield a family of mRNAs each encoding a different MSF protein. Exons	
CC	V and VI are thought to be related to the activity of the factor and	
CC	are also implicated in the stability, folding and processing of the	
CC	molecule. These exons are also thought to play a role in the observed	
CC	synergy of MSF with other cytokines. Exons V - XII are believed to be	
CC	implicated in the processing or folding of the appropriate structure of	
CC	the resulting factor, ie. one or more of these exons may contain	
CC	sequences which direct extracellular cleavage, adhesion, organisation of	
CC	the cellular matrix or extracellular matrix processing. Both naturally	
CC	occurring and non-naturally occurring MSF's may be characterised by	
CC	various combinations of alternatively spliced exons from this sequence,	
CC	with the exons spliced together in differing orders to form different	
CC	members of the MSF family.	
XX		
SO		
Sequence	1404 AA:	
Oy	Query Match	98.5%: Score 6470.5; DB 13; Length 1404;
Db	Best Local Similarity	87.5%; Pred. No. 0;
Matches	1229; Conservative	0; Mismatches 0; Indels 175; Gaps
Oy	1 MAWKTLPIYLLLSLVFIQOVSS-----	24
Db		
1	MAWKTLPIYLLLSLVFIQOVSS-----	24
Oy	25 -----	24
Db	61 kvvtaelackgcfcsferrgcdadqckkycpcpydfcaevhnptspsskap	120
Oy	25 -----	24
Db	121 ppsgaegtktstkrspkpknkkcttkvleseeitteesvsenqessssssssstlw	180

QY	25	-----VKDNKNRFRKKRPKPPVVDVAGSSGJDNDCDFKVTTPDST	6
Db	101	kikseksaaanreJqkklkvvdkhknrcfkkkprkpvvdeagsgjdngdfkvtlcpdst	2
QY	66	TOHNKVSYPKITPAKPINRPSLPSPDSRSTSTJWKKETTVETKETTNNKOSTDG	1
Db	241	tqhnkvstspkitlakinprpslppndstscscltvnkeetvctcttnqstsdg	3
QY	126	KEKTSKKEPOSTEKSADLAFTSVLAKPMPKAETTTKGALTPKEEPPMPKBPAS	1
Db	301	kektскеqsgtleksaklaptskvnlakrpkaeattkgalrtkreprrtkpkeas	3
QY	186	TTPEEPTPTTKSAPTTPEEPAPTTKKSAPTTKPKADPTTKKEPAPTTPKBPATTTKEP	2
Db	361	tkrkprrtkkapepttkrepaprrtkkapepttkkapepttkkapepttkkapepttkk	4
QY	246	APPTTKSAPTTPEEPAPTTKKBPATTKKEPAPTTKESPPPTPKBPAPTTKEPAPTTPK	3
Db	421	aprrtkseaprrtkrepaprrtkkapepttkkapepttkkapepttkkapepttkk	4
QY	306	EPAPTPKBPAPTTPEEPAPTTKBPADPTTKESPTPKBPAPTTTKSAPTTKKEPAP	3
Db	481	eprrtkkapepttkkapepttkkapepttkkapepttkkapepttkkapepttkk	54
QY	366	TKKSAPTTKESPTTKKEPAPTPPKBPAPTTPKKAPPTTKKEPAPTTKEPAPTTTKRP	42
Db	541	tkkseaprrtkreprrtkkapepttkkapepttkkapepttkkapepttkkapepttkk	60
QY	426	APTKKEPAPTTPKETAPTTPKLTPPTPKLAPTPPEKAPPTPEELPTTPEEPTPTT	48
Db	601	aprrtkrepaprrtkelaprrtkklrtkpekklaprrtkrepaprrtkelaprrtkrep	66
QY	486	PEEPAPTTPKAAANTPEEPAPTTPKBPAPTTKEPAPTTPKETAPTTKGAPTTKLP	54
Db	661	peepaprrtkkaaprrtkrepaprrtkkapepttkkapepttkkapepttkkapepttk	72
QY	546	APTPPKKAPKELAAPTTTKKPTSTTSKRAPPTPKGTAPTPKEBPAPTTPKBPAPTTK	60
Db	721	aprrtkkapepttkkapepttkkapepttkkapepttkkapepttkkapepttkk	78
QY	606	TAPPTTKKEPAPTTPKKAPKELAPTTKKPTSTTSKRAPPTPKGTAPTPPKBPAPTTPK	66
Db	781	taprrtkkapepttkkapepttkkapepttkkapepttkkapepttkkapepttkk	84
QY	666	KPAPTTPPEPPTPTSEVSPPTTKKBPPTTKKSDSTBELSAPPKALENSPKPEGPT	72
Db	841	kpaaprrtkprrtksevsrrtkkapepttkkapepttkkapepttkkapepttkk	90
QY	726	TKTPAPTKPEMTTAAOKTTBEDLTPPEPTTAAFKMKETATTTTEKTSKITAATTOY	78
Db	901	tkrpaaprrtkkapepttkkapepttkkapepttkkapepttkkapepttkkapepttk	96
QY	786	TSTTTTODTTPFKITTLKTTTTLAPKTTTKKTIITTELMNKPPEELAKKADATNTSKATTPK	84
Db	961	tsttttdttrfkittlktlkttlaprrtkkapepttkkapepttkkapepttkkapepttk	102
QY	846	POKPTPAKPKPSTKPKKPMRVKPKTTPTPKMTSTYMBELNPTSRIAEAMJOTTRPN	90
Db	1021	pokrrtkkapepttkkapepttkkapepttkkapepttkkapepttkkapepttkk	108
QY	906	QTPNSKLVENPKSEDAGAGETPHMLLPHEFMPVPTPDMOYLRRVNOGIIINPMUS	96
Db	1081	qtpnslkvevnpksedagagaeetphmlrrpvfmpvevcpdmoylrrvnoqiiinpmus	114
QY	966	DETNTCNCKPVDGTLTLTRNGIYAFKRYFWMLSPSPSPAPARITTEVWGISPIIDYFT	102
Db	1141	detntcnckpvdgtltrlnglvalfgyhfwmlspspspaparritlevngispiidvft	120
QY	1026	RCNBEKTFPFKOSQYWRFTNDIKDGYPRPIKGGJGGLGQIYVALSTAKYKKNPSEVY	108
Db	1201	rcnbeqrrtkffkdsqywrftndikdgyprpifkgygltsqglvalstakyrknpsevy	126

Run on: April 26, 2002, 16:16:50 ; Search time 138.84 Seconds
(without alignments)

Title: AA1
Perfect score: 6568
Sequence: 1 MAWKLPILLLLSVFVIO.....ARATTRSGQOTLSKVMVNC P 1229

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

```
Searched:      522463 seqs, 74073290 residues
Total number of hits satisfying chosen parameters: 522463
```

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```

Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

```

Database :

1: A.Genesecp_1101: *
2: /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT: *
3: /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT: *
4: /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT: *
5: /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT: *
6: /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT: *
7: /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT: *
8: /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT: *
9: /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT: *
10: /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT: *
11: /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT: *
12: /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT: *
13: /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT: *
14: /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT: *
15: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT: *
16: /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT: *
17: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT: *
18: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT: *
19: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT: *
20: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT: *
21: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT: *
22: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT: *

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	6470.5	98.5	1404	13	AAR26049	MSF precursor. SY
2	6470.5	98.5	1404	22	AAB60568	Human megakaryocyt
3	6470.5	98.5	1404	22	AAB29773	Human megakaryocyt
4	5894.5	89.7	1299	22	AAM23732	Human EST encoded
5	3484	53.0	902	22	AAB29778	Human MSF-derived
6	1707.5	26.0	472	22	AAB60569	Bovine MSF ortholo
7	1397.5	21.3	452	16	AAR80041	Human megakaryocyt
8	1168	17.8	5179	22	AAM24516	Human MSF ortholo
9	981	14.5	763	21	AAAG38942	C899F predicted am
10	950	14.5	1664	19	AAAG43106	Arabidopsis thalia
11	772.5	11.8	1325	22	AAW03645	C. thermocellum O1
						Peptide #237 enco

12	715	10.9	763	18	AAW31852
13	625.5	9.5	4412	21	AAV53666
14	5612	9.3	572	18	AAW31855
15	553.5	8.4	844	7	AA605070
16	542	8.3	807	21	AAV54467
17	520.5	7.9	788	21	AAV54466
18	508	7.7	1837	21	AAW11726
19	506.5	7.7	744	9	AA682975
20	486.5	7.6	2971	21	AAW41231
21	489.5	7.5	2972	22	AAW50363
22	489.5	7.5	3118	22	AAW50362
23	489	7.4	1721	21	AAW11727
24	488.5	7.4	826	13	AAW26042
25	488	7.4	617	22	AAW16458
26	488	7.4	617	22	AAW04187
27	488	7.4	957	21	AAV59288
28	488	7.4	957	22	AAW24513
29	488	7.4	1721	19	AAW48299
30	485	7.4	1127	22	AAW5541
31	476.5	7.3	652	9	AA682974
32	467.5	7.1	511	22	AAW14893
33	467.5	7.1	511	22	AAW27312
34	467.5	7.1	511	22	AAW02607
35	450.5	6.9	378	12	AAW4160
36	446.5	6.8	378	12	AAW41462
37	446.5	6.8	750	20	AAV05477
38	443	6.7	2870	21	AAV95559
39	442.5	6.7	3178	21	AAV95556
40	442.5	6.7	751	16	AAW80839
41	435.5	6.6	2819	22	AAW35408
42	431.5	6.6	2665	22	AAW4533
43	431.5	6.6	2665	22	AAW6950
44	431.5	6.6	2665	22	AAW02259
45	431.5	6.6	3266	21	AAW42491

ALIGNMENTS

RESULT	1
AA26049	standard; protein; 1404 AA.
ID	AA26049
XX	AA26049;
AC	
XX	02-FEB-1993 (first entry)
DT	
XX	
DE	MSF precursor.
XX	
KW	Megakaryocyte colony stimulating factor; secretion signal; meg-CSF stability; proteolytic cleavage; adhesion; alternative splicing.
XX	
OS	Synthetic.
XX	
FH	Key
FT	Location/Qualifiers
FT	1..26
FT	/label= Exon_I
FT	26..67
FT	/label= Exon_II
FT	67..107
FT	/label= Exon_III
FT	107..157
FT	/label= Exon_IV
FT	157..200
FT	/label= Exon_V
FT	200..1141
FT	/label= Exon_VI
FT	1141..1166
FT	/label= Exon_VII
FT	1166..1212
FT	/label= Exon_VIII
FT	1213..1266
FT	/label= Exon_IX
FT	

CC factor, ie. one or more of these exons may contain sequences which
 CC direct proteolytic cleavage, adhesion, organisation of the cellular
 CC matrix or extracellular matrix processing. Both naturally occurring
 CC and non-naturally occurring MSF's may be characterised by various
 CC combinations of alternatively spliced exons from this sequence, with
 CC the exons spliced together in differing orders to form different
 CC members of the MSF family.

XX Sequence 111 AA;

Query Match 8.4%; Score 613; DB 13; Length 111;
 Best Local Similarity 100.0%; Pred. No. 8.1e-28;
 Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 LSSCAGRCGEGYSRDPATCNCQYHMECCPDFFKRYCTAELSCKGRCESPFERGECDC 86
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 8 lsscsgrcgegyrsdctncdncqhyneccpdkrvctaelsckgrcfesfergrecdc 67
 QY 87 DAQCKKIDKCCPDIESFCAEYHNPSTSPSSKKAAPPSSGASQTIK 130
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 68 daqckkydkccpdyesfcaevhnpstspsskkaappssgasqtlk 111

Search completed: April 26, 2002, 16:25:34
 Job time: 524 sec

PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134270.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140655.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140981.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144332.
PR 20-JUL-1999; 99US-0144332.
PR 21-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145293.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147322.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148341.
PR 16-AUG-1999; 99US-0148341.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149375.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 25-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154479.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0156359.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 18-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.

(e.g. by enzyme linked immunosorbant assay (ELISA)). AA128460 to AA129512 CC and AAM24494 to AAM24523 represent nucleotide and amino acid sequences CC given in the exemplification of the present invention.

XX Sequence 5179 AA:

Query Match 16.3%; Score 1194; DB 22; Length 5179;
Best local similarity 30.6%; Pred. No. 9,4e-59;
Matches 400; Conservative 74; Mismatches 515; Indels 320; Gaps 51;

QY 4 KTLPIYILLLSVFIQOVSODLSSC--AGRCG-----EGYSRDATCNDYMCQHMEC 56
DB 1169 ktlpily-----eodlkcvctadkcgcyedhhyppgsaspreetckascv-c 1213
QY 57 CPDFKRYCTAELSCKRCFESFERG---RECCDAQCKKYDKC-----PRYESFCAE 106
DB 1214 tnsqgvrcrpe---egklngtdgafcywelcgmgtvekhfnfcsitrlptltltft 1270
QY 107 VHNFTSPS-SKRAPPSGASORTIKSTTKRSPPKPKKKTKVIESPEITEVDKNKKNRT 165
DB 1271 itlptlslstltltltltlptslstl---pk-----lccldwadine----- 1311
QY 166 KKKPTPKPVVDAGSGIDNGDFK-----VTPPDSTTGH-NKVSISPK 208
DB 1312 -----dhpsgsddgdrepfdgvcgapedlccrsvkdphtlsleghgqkvqcavs 1360
QY 209 -----TTAKPINRPSLPPNSDKETSL 233
DB 1361 vglcknedqfngpfglcydklrvncowpmdkcltspptltspptltltlpt 1420
QY 234 TVNKEETVETETTTNKQITSDGKEKTSAKETQSIKTSAKDLAPTSKVLAKPTKAE 293
DB 1421 tpspplt-----tttpttltspptlttpt-----lpt-----tpsppls 1459
QY 294 TTTKAPALTREKPT-----PTTKEPASTPKKEPTPTIKSAP--TTPKPE-----AP 340
DB 1460 tttpttltspptltspptltspptlttpttlttpttlttpttlttpttlttpt 1519
QY 341 TTTKAP--TTPEAPATTKEPAPTTPKEP-----APTTEKAPATTKSAP--TTPK 390
DB 1520 tttspptlttpttlttpttlttpttlttpttlttpttlttpttlttpttlttpt 1579
QY 391 EPAPTTPKKAPATTPKKAPATTPKKAPATTPTTKEPAPTTPKKEPAPTTPKPKP-----APTA 443
DB 1580 pttlttspptlttpttlttpttlttpttlttpttlttpttlttpttlttpt 1639
QY 444 PKKPAPTTPKEPAPTTPKKEPAPTTPKKEPAPTTPKAPATTTKSAP 503
DB 1640 tpsppltlttpttlttpttlttspptlttpttlttpttlttpttlttpttltmtp 1699
QY 504 ---TTPKEPSTTPTEKAPATTKEPAPTTPKKAAPTTPKKEPAPTTPKKAAPT 560
DB 1700 spptltspptlttpttlttpttlttpttlttpttlttpttlttpttlttpt 1758
QY 561 AKKEPAPTTP-----KEPAPTTP----- 578
DB 1759 tlpplslpttptlspstlttpttccvplcnvgtvldsgkpnfhkpggdclldgvcgpgwa 1818
QY 579 -----KKLTP----- 583
DB 1819 anisctatnypdipiglgvtvcdvsvglcknedqkpgvlpmafclnyelnvqccoc 1878
QY 584 -TTEKLAPTTPKEPAPTTPPEELAPTTPEEPPTT--TPEEPAPTTPKAAPNTPKEPAPT 641
DB 1879 vtqptmtltttenppttlttlttvtprpctgctqtpptlttlttvtprpctg 1938
QY 642 PKEP-----APTTPKEPAPT--TPEKAPATTPKGTAPTTIKKAPATTPKKAPKELAPTT 695
DB 1939 tqpptlttlttvtprpctgctqtpptlttlttvtprpctgctqtpptlttltt 1998
QY 696 KEPTSTTSKAP--TTPKGTAPTTPEKAPATTTPKKEPAPTTPKGTAPTTIKKAPATTPKKP 754

DB 1999 vtprpctgctqtpptlttlttvtprpctgctqtp--tpttlttvtprpctgctqtp 2057
QY 755 APKELAPTTTGPSTSDKAP--TTPKEAPTTPKKAPATTTPKKAAPTTPPTTSE 813
DB 2058 ttpptlttvtprpctgctqtpptlttlttvtprpctgctqtp--tpttlttvtpr 2116
QY 814 VSTPTTEPTTHKSPDESTPELSAEPKALENSKPE-----GVPT-TKTPA 862
DB 2117 tprpctgctqtp--tpttlttvtprpctgctqtpptlttlttvtprpctgctqtp 2173
QY 863 AKPEMTTAKDKTERDLR--TTP--ETTTAAPKMT--KETATTEKTESKITATT 914
DB 2174 tpttlttvtprpctgctqtpptlttlttvtprpctgctqtpptlttlttvtpr 2233
QY 915 TQVSTTTQDTTPFKITTLKTTTTLAPKVTTP--KTIITTELMNNPEEKAKPDATNSKA 973
DB 2234 tp-tgtppttpt--ltt--lttvtprpctgctqtpptlttlttvtprpctgctqtp 2288
QY 974 TTPKQKPTKAP--KKPTSTKKKPTMPRVKPKPTPTPKMTSTMPELNP---TSRIAEAM 1029
DB 2289 tpttlttvtprpctgctqtpptlttlttvtprpctgctqtpptlttlttvtpr 2348
QY 1030 LQTTTP-PNQTTPNSKLVEVNPKSEDAAGAEETPMMLLRPHVFMPEVTP 1077
DB 2349 tprpctgctqtpptlttlttvtprpctgctqtp--lqtp--tpttlttvtpr 2392

RESULT 9
AAG38942
ID AAG38942 standard; Protein: 763 AA.
XX
AC AAG38942;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 48115.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
PD
XX
XX 06-SEP-2000.
PF
XX 25-FEB-2000; 2000EP-0301439.
PR
XX 25-FEB-1999; 990S-0121825.
PR 05-MAR-1999; 990S-0123180.
PR 09-MAR-1999; 990S-0123548.
PR 23-MAR-1999; 990S-0125788.
PR 25-MAR-1999; 990S-0126264.
PR 29-MAR-1999; 990S-0126785.
PR 01-APR-1999; 990S-0127462.
PR 06-APR-1999; 990S-0128234.
PR 08-APR-1999; 990S-0128714.
PR 16-APR-1999; 990S-0129845.
PR 19-APR-1999; 990S-0130077.
PR 21-APR-1999; 990S-0130449.
PR 23-APR-1999; 990S-0130510.
PR 23-APR-1999; 990S-0130891.
PR 23-APR-1999; 990S-0131449.
PR 28-APR-1999; 990S-0132048.
PR 30-APR-1999; 990S-0132407.
PR 04-MAY-1999; 990S-0132484.
PR 05-MAY-1999; 990S-0132485.
PR 06-MAY-1999; 990S-0132486.
PR 06-MAY-1999; 990S-0132487.
PR 07-MAY-1999; 990S-0132863.
PR 11-MAY-1999; 990S-0134236.
PR 14-MAY-1999; 990S-0134218.

```

XX DE Bovine MSF orthologue, superficial zone protein (SZP).
XX XX
XX KM Bovine: CACP protein; campodactylly-arthropathy-coxa vara-pericarditis;
XX KM superficial zone protein; SZP; MSF orthologue; synovial lubricant;
XX KM osteoarthritis; joint lubrication; osteopathic; antiarthritic.
XX OS Bos taurus.
XX PN WO200107068-A1.
XX PD 01-FEB-2001.
XX XX
XX PF 21-JUL-2000; 2000MO-US20002.
XX PR 23-JUL-1999; 99US-0145328.
XX PR 19-JUL-2000; 2000US-0145328.
XX PA (UYCA-) UNIV CASE WESTERN RESERVE.
XX XX
XX PI Warman ML;
XX DR WPI; 2001-182721/18.
XX PT New composition comprising the campodactylly-arthropathy-coxa
XX PT vara-pericarditis protein in combination with an anesthetic, useful for
XX PT treating osteoarthritis, or as lubricants of tissue and joints
XX PS Example 1; Fig 4; 34pp; English.
XX XX
XX CC The invention relates to a method of treating osteoarthritis via the
XX CC administration of a composition comprising the campodactylly-arthropathy-
XX CC coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.
XX CC The composition may further comprise a local anesthetic. The composition
XX CC of the invention may be administered via intra-articular or intravenous
XX CC injection. The human CACP protein is identified in the invention as
XX CC being megakaryocyte stimulating factor (MSF). The gene encoding
XX CC CACP protein (MSF) is located on chromosome 1q25-31 and mutations in
XX CC this gene are responsible for the heritable disorder campodactylly-
XX CC arthropathy-coxa vara-pericarditis, in which patients have synovial
XX CC hyperplasia without evidence of inflammation. CACP protein (MSF)
XX CC acts as a synovium lubricant, and can be used to lubricate tissue and
XX CC joints in the treatment of osteoarthritis. The composition may be
XX CC applied to reduce the symptoms of osteoarthritis (e.g., joint pain,
XX CC loss of range of movement or joint damage). The present sequence
XX CC represents the bovine orthologue of human MSF, superficial zone
XX CC protein (SZP).
XX SQ Sequence 472 AA;

```

```

Db 314 a1slgqysrpesvyffkrgsvqgylykgeptqkclgrpalnysvgetagvrrrrfe 373
Qy 1263 RAIGSGYHTIRIOYSPARLAYODKGVLANEKKVSIILMRGJPNVWTSISIPNIRKPDGy 1322
Db 374 raigsgyhtlrlhytprvpygdkgflhnekvkstlwrjpnvvtalsipnlrkpdgy 433
Qy 1323 DYAFSKDOYYNIDVPSTARAITTRSGQTLISKWYNCP 1361
Db 434 dyafskdqyynldvpstaraaitrsgqlslntwncp 472

RESULT 8
AAM24516
ID AAM24516 standard; Protein: 5179 AA.
AC AAM24516;
XX XX
XX DT 12-OCT-2001 (first entry)
XX XX
XX DE C899P predicted amino acid sequence.
XX XX
XX KW Human; immunotherapy; diagnosis; colon cancer; colon tumour;
XX KW immunogenic; gene therapy; vaccine; colonic cancer.
XX OS Homo sapiens.
XX PN WO200149716-A2.
XX PD 12-JUL-2001.
XX PF 29-DEC-2000; 2000MO-US35596.
XX PR 30-DEC-1999; 99US-0476296.
XX PR 10-JAN-2000; 2000US-0480321.
XX PR 15-FEB-2000; 2000US-0504629.
XX PR 06-MAR-2000; 2000US-0519444.
XX PR 19-MAY-2000; 2000US-0575251.
XX PR 29-JUN-2000; 2000US-0609448.
XX PR 28-AUG-2000; 2000US-0649811.
XX PA (CORI-) CORIXA CORP.
XX PI Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
XX PI King GE, Wang T, Jiang Y;
XX DR WPI; 2001-441847/47.
XX XX
XX PT Colon tumor associated proteins and nucleic acids useful for the
XX PT prevention, diagnosis and treatment of colonic cancer -
XX PS Claim 2: Page 446-462; 472pp; English.
XX XX
XX CC The present invention describes colon tumour associated proteins (I) and
XX CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
XX CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
XX CC (II) may be used in the prevention, diagnosis and treatment of diseases
XX CC associated with inappropriate colon tumour associated protein (TCAP)
XX CC expression, such as colonic cancer. For example, (I) and (II) may be
XX CC used to treat disorders associated with decreased expression by
XX CC rectifying mutations or deletions in a patient's genome that affect the
XX CC activity of TCAPs by expressing inactive proteins or to supplement the
XX CC patients own production of them. Additionally, (II) may be used to
XX CC produce the TCAP proteins, by inserting the nucleic acids into a host
XX CC cell culturing the cell to express the protein. (II) and its
XX CC complementary sequences may also be used as DNA probes in diagnostic
XX CC polymerase chain reaction (PCR) and hybridisation assays to detect and
XX CC quantitate the presence of similar nucleic acids in samples, and
XX CC therefore which patients may be in need of restorative therapy. (I) may
XX CC also be used as antigens in the production of antibodies against TCAPs
XX CC and in assays to identify modulators of TCAP expression and activity.
XX CC Anti-(I) antibodies and antagonists may also be used to down regulate
XX CC TCAP expression and activity. The anti-(II) antibodies may also be used
XX CC as diagnostic agents for detecting the presence of TCAPs in samples

```



```
QY 978 POKPKAKPKSTTKKPKMPVRKPKTPTPKMTSTMPKELNPTSTIAEMLQTTTRPN 1037
DB 1021 pdkpckapkkpckckkpkmpvrkpktpckrckmtstmpelnpckrckaeamqcttrpn 1080
QY 1038 QTPSKLVEVNPKESEDGSGEGEPHMLLRPHVPMPEVTTDDMDYLPKRVNOGIINPMJLS 1097
DB 1081 qtpmsklvevnpksegdsgaegeclpmllrphvfmpcvtrpdmxylpvrpqgqlllnpmls 1140
QY 1098 DETNICKGKPEVDGLTTLRNGTLVAFRGHYPMMLSPFSPSPARTRIEWGIPSPIDTPT 1157
DB 1141 detlnckgkpvdgllttrngtlvafrghyfwmllspfspspartriewgipspidvlt 1200
QY 1158 RCMCEGTEFFEKDSQVRYRPNNDKDGKYPKPIKGFGLGTGIVAAALSTAKKKNPESVY 1217
DB 1201 rcmcegtcltffkdsqywrftndkldgypkplkfgslgtqivaalstakknwpsvay 1260
QY 1218 FFRGSGSIQOYIYKQEPVQKCPGRPALNVPYGEVTOVRRRFEALPSSQTHTRIOY 1277
DB 1261 ffrgsgsiqyiykqepvqkcpgrpalnvpvygevtvrrrrferraldpsqthtrlqy 1320
QY 1278 SPARLAYOOKGVLAHNEKYSILMRGLPNVVTSAISLPNIRKPDGYDYAFSKDQYXNIDY 1337
DB 1321 sparlaysqdgylahnevksilwrglpnvvtalsipnirkpdydyafskdyqynldv 1380
QY 1338 PSRTARAITTRSGQTLSTKYWYNCP 1361
DB 1381 psrtaraittrsgqlstkywyncp 1404

RESULT 3
AAB29773
ID AAB29773 standard; Protein: 1404 AA.
AC AAB29773;
XX
XX
XX 28-FEB-2001 (first entry)
DE Human megakaryocyte stimulating factor (MSF), SEQ ID NO.1.
XX
XX
XX Human MSF: megakaryocyte stimulating factor: tribonectin;
XX alternative splicing; joint boundary lubricant; O-linked oligosaccharide;
XX osteoarthritis; tribosupplementation; tissue adhesion inhibition;
XX friction coefficient reduction; gene therapy; antiarthritic;
XX osteoponthic.
XX
XX Homo sapiens.
OS
XX
XX MO200064930-A2.
XX
XX 02-NOV-2000.
XX
XX 24-APR-2000; 2000MO-US10953.
XX
XX 23-APR-1999; 99US-0298970.
XX
XX (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.
XX
XX Jay GD;
XX
XX WPI: 2001-024673/03.
XX
XX N-PSDB: AAC81498.
XX
XX Novel tribonectin polypeptide useful as lubricant for treating
XX osteoarthritis, comprises O-linked lubricating moiety
XX
XX Claim 3; Page 7; 47pp; English.
XX
XX The invention relates to a human tribonectin which is a product of
XX alternative splicing of the human MSF (megakaryocyte stimulating factor)
XX gene. The tribonectin has at least one O-linked oligosaccharide
XX lubricating moiety and has a polypeptide sequence comprising 1-76
XX repeats of a motif having at least 50% identity to the sequence KKPAPPT
XX (AAB29774). The invention also relates to a nucleic acid encoding a
```

```
CC human MSF-derived tribonectin; a biocompatible composition comprising a
CC human tribonectin for inhibiting tissue adhesion formation; and a method
CC of diagnosing osteoarthritis or a predisposition to osteoarthritis by
CC measuring the amount of MSF or its fragment in a biological sample of a
CC mammal, wherein an increased amount of MSF compared to a control
CC indicates the presence of or predisposition to developing
CC osteoarthritis. The tribonectin and DNA encoding it are useful in the
CC treatment of osteoarthritis, where they may be used for lubricating
CC mammalian joints, such as articulating joints of humans, dogs or horses.
CC The tribonectin, when formulated as a membrane, foam, gel or fibre, is
CC useful for inhibiting adhesion between two surfaces such as the injured
CC or trauma of a mammal, where the injury is caused by a surgical insertion
CC of a particular, one of the surfaces is pericardial tissue. DNA encoding a
CC tribonectin may be used in gene therapy. The present sequence represents
CC human MSF.
XX
XX
XX Sequence 1404 AA:
SQ
```

```
Query Match 99.6%; Score 7291.5; DB 22; Length 1404;
Best Local Similarity 96.9%; Pred. No. 0;
Matches 1361; Conservative 0; Mismatches 0; Indels 43; Gaps 1;
```

```
QY 1 MAMKTLPIYLLLSVFIQOVSQDLSSCAGRCGEYSKDATCNDYNCQHYMECCPDF 60
DB 1 mamktlpiyllllsvfivqvsqdlsscagrcgeysrdatcncdncqhyameccpdf 60
QY 61 KRVCTAELSCGRCFESPERGECDDACCKKYDKCCPDYESCAYEHNTSPSSKKAP 120
DB 61 krvtaelscgrcfespergrecddacckkydkccpdyescayehntspsskkap 120
QY 121 PPSGASQTIKSTTKRSPKPPNKKKKVIESEITE----- 156
DB 121 ppsgaqqliksttkrskppnkkkkvleeseeelhevsengessssssssstllw 180
QY 157 -----VKDNKKNRTRKKRPKPPVYVDGSGLDNDFVYTPDST 197
DB 181 kiksksnaaanrlqkklkylkdknrtkkpdkpvdgsgldngdfvyltclpdtat 240
QY 198 TQNNKSTSKITTYAKPINRPSLPNSDTSKETSLTVNKETVETKTTNNKQSTG 257
DB 241 tqnnkstskltitakpindrpslpsndtsketskltvnnkettvettcnkqstldg 300
QY 258 KEKTSAKETOSIEKTSANDLAPTSVLAKPPPKATETTKGALTYTPKEPTPKEPAS 317
DB 301 kektsaketosiektsakdlaptsvklakppkaetttkgalttptkcpckepas 360
QY 318 TTPKEPTPTTIKSAPTPKEPAATTTKASAPTPKKEPAATTTKEPAATTTKEP 377
DB 361 ttpkeptpttiksaptpkepaatttkaspptpkkepaatttkepaatttkepa 420
QY 378 APTTKASAPTTKEPAATTPPKKPAATTTKEPAATTTKEPAPTTPK 437
DB 421 apttkasaptpkepaattpkkpaatttkepaatttkepaattpk 480
QY 438 EPAPTAAPKKAAPTTKEPAATTPKKEPAPTTTKEPSPTPKKEPAATTTK 497
DB 481 epaptapkkaptpkkepaattpkkepaattpkkesptpkkepaatttk 540
QY 498 TTKSAPTTPKESPPTTTKEPAATTPKKEPAATTPPKKPAATTTKEPAATTTK 557
DB 541 ttksaptpkkespptttkepaattpkkepaattpkkpaatttkepaattpk 600
QY 558 APTAPKEPAATTPKETAATTPPKKLPTTPPEKALPTTPKKAAPTTPEEAPPT 617
DB 601 aptapkepaattpketaattpkkllpttppekaltptpkkaapttpeeappt 660
QY 618 PEEAPPTPKKAAPVTKEPAATTPKKEPAATTPKKEPAATTPPKGATTTTKEP 677
DB 661 peeapptpkkaapvtkpepaattpkkepaattpkkepaattpkgattpckep 720
QY 678 APTTPPKKAPKELAPTTTKEPTSTTSOKPAATTTKGAATTPKKEPAATTTPKG 737
```


OY 1218 FFKRGSIQIYYIKQEPVQCPCRRPALNIVYGEMTVORRRRRERARLGPSCGHTIRQY 1277
| | | | |
Db 1261 ffkrgsqidqiykygepqkcpgrrrpalnlypygemtvvrrrrrteralgsqchrlirgdy 1320

OY 1278 SPARAYODKCVLNHEVSVSLIMRGCLPWWVSATSLIPRIKRKPDGDYAFSKDOYYNIDV 1337
| | | | |
Db 1321 sparaydkqyvlmhevsvslwrglphvwvsalslpnlrrpdqdyafskdgyndlv 1380

OY 1338 PSRTARAITTRSGQTLASKWYNCP 1361
| | | | |
Db 1381 psrtaraittrsqtlskwynncp 1404

RESULT 2
AAB60568
ID AAB60568 standard; Protein; 1404 AA.

XX AAB60568;
AC
XX
DT 27-APR-2001 (first entry)
XX
DE Human megakaryocyte stimulating factor (MSF, CACP).
XX
KW Human; CACP protein; campodactylly-arthropathy-coxa vara-pericarditis;
KM MSF, megakaryocyte stimulating factor; synovial lubricant;
KW Chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathlc;
KW antiarthritic.
XX
OS Homo sapiens.
XX
PN WO200107068-A1.
XX
PD 01-FEB-2001.
XX
PF 21-JUL-2000; 200OWO-US20002.
XX
PR 23-JUL-1999; 990US-0145328.
XX
PR 19-JUL-2000; 200OUS-0145328.
PA (UYCA-) UNIV CASE WESTERN RESERVE.
PI Warman ML;
XX
DR WPI; 2001-182721/18.
XX
PS
XX

Example 1; Page -: 34pp; English.

The invention relates to a method of treating osteoarthritis via the administration of a composition comprising the campodactylly-arthropathy-coxa vara-pericarditis (CACP) protein, or portions of the CACP protein. The composition may further comprise a local anaesthetic. The composition of the invention may be administered via intra-articular or intravenous injection. The human CACP protein is identified in the invention as being megakaryocyte stimulating factor (MSF). The gene encoding CACP protein (MSF) is located on chromosome 1q25-31, and mutations in this gene are responsible for the heritable disorder campodactylly-arthropathy-coxa vara-pericarditis, in which patients have synovial hyperplasia without evidence of inflammation. CACP protein (MSF) acts as a synovium lubricant, and can be used to lubricate tissue and joints in the treatment of osteoarthritis. The composition may be applied to reduce the symptoms of osteoarthritis (e.g., joint pain, loss of range of movement or joint damage). The present sequence represents human megakaryocyte stimulating factor (MSF, CACP protein). Note: This sequence is not given in its entirety in figure 4 of the specification, although a Genbank accession number was given. This sequence was therefore obtained from Genbank (U70316).

SQ Sequence 1404 AA;

	Query Match	99.6%;	Score 7291.5;	DB 22;	Length 1404;	
	Best Local Similarity	96.9%;	Pred. No. 0;	Mismatches	Indels	Gaps
	Matches 1361;	Conservative	0;			
Oy	1 MAMTLPITYLLILLSFYIQOVSODLSSCAGRCGEYSRDATNCNDYNOCHMYECCPDF	60				
Dd	1 mawtlrplyllllysvrlqysgdlscagrcgeysrdatncndynohymecpdf	60				
Oy	KRYVTAELSCGRGCFESFERREDCDAOCKKYDKCCPDYSPFAEVHNTPSPSSKKAP	120				
Dd	kryvtaelscgrgcfeesferreceddaockkydkccpdysesfaevhnlpssskkap	120				
Oy	PPSASQTIKSTTKRSRKPNNKKTKKVSESEITE-----	156				
Dd	ppsasqlkstktrspkkaktkkvseeseicehmsvengeassssssstslw	180				
Oy	121 ppsasqlkstktrspkkaktkkvseeseicehmsvengeassssssstslw	180				
Dd	121 ppsasqlkstktrspkkaktkkvseeseicehmsvengeassssssstslw	180				
Oy	157 -----VKDNKNKRTKKRPKPVWDEAGSLDNGDFKVTPTPTST	197				
Dd	151 kksksnsaanrelqklkvdknkrkttkpprvpvdeagsldngdfkvtptst	240				
Oy	198 TOHNVKSTSPKITAKINRPSIPNSDTSKETSLTVNKEHTVETKTPTTNKOITSDG	257				
Dd	241 tqmhvsfspbiltakinpripspnadskeetslvnkeltvetkcettlnqtstdg	300				
Oy	256 KEKTSASETOSIEKTSBKDLAFTSKYLAKPPTKAETTGGALTTPEKBPPTTPKEPAS	317				
Dd	301 kektsakeitgsiektsackdlaptskylakprpkkaettkgalttpekppttpkepasp	360				
Oy	318 TPPEKPPPTTKSAPTTPKEPAPTTKKSAPTTPKEPAPTTKKEPAPTTKKEPAPTTKEP	377				
Dd	361 ttpkeprrpttklsapttpekpafttkksapttksepaplttksepaplttksepaplttksep	420				
Oy	378 APPTTKSAPTPKKEBAPTTPKKAPATTPKPBABTTPKEPPTTPKBPAPTTPKEBAPTTPK	437				
Dd	421 appttksaptpkbpaptpirkpapattpkpbabtpirkperrtpirkbpaptpirkbpaptpir	480				
Oy	438 EPAPPAKKKPAFTTPKEBAPTTPKEBAPTTPKESPPTTPKEBAPTTPKSAPTTPKEBAPT	497				
Dd	481 epaptkappkrapltirkbpapltirkbpapltirkespirttksepaplttksaplttksepap	540				
Oy	498 TTKSAPTTPKESPPTTPKEBAPTTPKKEBAPTTPKKAPATTPKPBABTTPKBEAPATTPKKP	557				
Dd	541 tkfsaptpirkespirttksepapltirkbpaptpirkpapattpirkbpaptpirkbpapltik	600				
Oy	558 APAPAKEBAPTTPKETAAPTTPKLLTPPYTEKLATTPEKAPPTTPBELAPTPPEBPTPT	617				
Dd	601 apapakepapttketapttklltpyrtekelattpekappttpbelaptpceepptc	660				
Oy	618 PEEBAPTTPKKAAPNTPKBPAPTTPKKEBAPTTPKEBAPTTPKETAAPTTPKGATTLKBP	677				
Dd	661 peebaptpckaapnrpkbpaptpirkbpaptpirkbpaptpirkbpaptpirkbpapltike	720				
Oy	678 APPTPKKPAPELKELAPTTTKEPTSTSDDKPAFTTPKGAATTPKKEBAPTTPKGPB	737				
Dd	721 aptpkkpapakelaarttkptsdsdpaplttkgaattpirkbpaptpirkbpaplttkg	780				
Oy	738 TAFTPLKEBAPTTPKKBAPELKELAPTTTGPTSTSDSKPAFTTPKETAAPTTPKKEBAPTTPK	797				
Dd	781 taftplkebabtpkkbpakelaarttkgptsdsdpaplttkgaattpirkbpaptpirkbpap	840				
Oy	798 KPAPTTPERPPTTSEYSTPTTKEPTTHHSDPDESIPBELSABETPALENSKEPGVP	857				
Dd	841 kpapttpeprtptseystpttkcptllhkspdesipelsaabelpalenskepgvp	900				
Oy	858 TKPAPAARKPEMTTAAADKTTENDLRTPPTTTAAFKMKEKATTTTEKTTESKIATATTQV	917				
Dd	901 tkpapaarkpemttadkkttenrlrtpttttaafkmkekattttekstektatltty	960				
Oy	918 TSTTQDTPEPFKTTTTLTTLAPKVVTTKKTTTTETIINNKPBEETAKPRDRATNSKATTP	977				
Dd	961 tstctgdtpfkltlkttltlapkvttkkittlteimnpbeetakprdratsnaktpk	1020				


```

FT      Region      1266..1331
FT      /label= Exon_X
FT      Region      1331..1373
FT      /label= Exon_XI
FT      Region      1373..1404
FT      /label= Exon_XII
XX      MO9213075-A.
XX      06-AUG-1992.
XX      17-JAN-1992: 92WO-US00433.
XX      18-JAN-1991: 91US-0643502.
XX      10-SEP-1991: 91US-0757022.
XX      (GENM ) GENETICS INST INC.
PI      Clark SC, Geener TG, Hewick RM, Jacobs K, Turner K;
XX      WPI: 1992-284660/34.
XX      N-PSDB: AAQ27223.
XX      New human mega-karyocyte stimulating factors - for treating
XX      immune deficiencies, cancer, exposure to radiation or drugs,
XX      bacterial and viral infections, etc.
XX      Claim 1, 2 and 3: Fig 1; 87p; English.
XX      The sequence given is a full length translation from the megakaryocyte
XX      stimulating factor (MSF) precursor. The sequence covered by exons II,
XX      III and IV encodes megakaryocyte stimulating factor (MSF). This
XX      sequence is modified by the addition of an N-terminal sequence encoding
XX      a secretory leader, an initiating methionine proceeding exon II and a
XX      terminating codon following exon IV. The cDNA sequence given contains
XX      (meg-CSF). Exon I contains the initiating methionine, and encodes a
XX      classical mammalian protein secretion signal sequence. The sequence
XX      encoding the original meg-CSF includes exons II-IV and is thought to
XX      terminate in the region between amino acid residues 134 - 147. The
XX      primary transcript of this gene may be cleaved in different ways to
XX      yield a family of mRNA's each encoding a different MSF protein. Exons
XX      V and VI are thought to be related to the activity of the factor and
XX      are also implicated in the stability, folding and processing of the
XX      molecule. These exons are also thought to play a role in the observed
XX      synergy of MSF with other cytokines. Exons V - XII are believed to be
XX      implicated in the processing or folding of the appropriate structure of
XX      the resulting factor, ie. one or more of these exons may contain
XX      sequences which direct proteolytic cleavage, adhesion, organisation of
XX      the cellular matrix or extracellular matrix processing. Both naturally
XX      occurring and non-naturally occurring MSF's may be characterised by
XX      various combinations of alternatively spliced exons from this sequence,
XX      with the exons spliced together in differing orders to form different
XX      members of the MSF family.
XX      Sequence 1404 AA:
XX
Query Match      99.6%: Score 7291.5; DB 13; Length 1404;
Best Local Similarity 96.9%: Pred. No. 0;
Matches 1361; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

```

```

QY      157 -----VKONKKRKRKKRPKPPVVDAGSLGNDGKVTTPDTST 197
DB      181 kksksksaanrejqkklkvkdnknrrkkkcpkpvpvdaagsglndgdfkvtcpts 240
QY      198 TOHNVSTSPKITTAKPINPSPSLPNSDTSKETSLVYVNRKETTVEETKETTNNKOTSPDG 257
DB      241 tqhkvstspklttakpnpnpslpnpsdstskelslvnketvclvtctcmqkscdg 300
QY      258 KEKTSAREKOSIETSKADLAPTSKVLAFTPKAETTTKGPALJTRKEPTTPPKEPAS 317
DB      301 kektsarektslektsakdlaptskylakpckpaetltkxpaljtrkepttpkpas 360
QY      318 TTPKEPTTPPKSAPTTPKEPAPTTPPKSAPTTPKEPAPTTPKEPAPTTPKEPA 377
DB      361 ttpkepttpksaptpkpaaptpkpaaptpkpaaptpkpaaptpkpaaptpkpaaptpk 420
QY      378 APTTTSAPTPPKPAPTTPPKPAPTTPPKPAPTTPPKPAPTTPPKPAPTTPPKPA 437
DB      421 apttksaptpkpaaptpkpaaptpkpaaptpkpaaptpkpaaptpkpaaptpkpa 480
QY      438 EPAPTAPEKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 497
DB      481 epaptapekpaaptpkpaaptpkpaaptpkpaaptpkpaaptpkpaaptpkpaap 540
QY      498 TTKSAPTPKEPSFTTKKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPA 557
DB      541 tksaptpkpsfttkkepaaptpkpaaptpkpaaptpkpaaptpkpaaptpkpaap 600
QY      558 APTAPKEPAPTTPKEPAPTTPPKKLTPTTPPKLAFTTPKEPAPTTPBELAFTTP 617
DB      601 aptapekpaaptpkpaaptpkpaaptpkpaaptpkpaaptpkpaaptpkpaaptp 660
QY      618 PEEBAPTPKAAAPNTPKEBAPTPKEBAPTPKEBAPTPKEBAPTPKEBAPTPKEB 677
DB      661 peebaptpkaaanptkebaaptpkebaaptpkebaaptpkebaaptpkebaaptpke 720
QY      678 APTPKPKPAKELAPTTTKEPTSTSDKAPPTPKGATTPKEBAPTPPKBAPTPPK 737
DB      721 aptpkpkpkelaptttkeptstsdkapptpkgattpkebaaptpkbaaptpkba 780
QY      738 TAPTTKEBAPTPPKKPAKELAPTTTSGPTSTSDKAPPTPKGATTPKEBAPTPPK 797
DB      781 tapttkebaaptpkkpakeaptppttsgptstsdkapptpkgattpkebaaptpk 840
QY      798 KPAPTPEPPTSTSVSTPTTKEPTTHKSPDESPELSAEPKALENSKEPEGYPT 857
DB      841 kpaptpepptstsvstpttkeptthkspdespeelsaepkalaenskepegypt 900
QY      858 TTPAATKPEMTTAKDKTTERDLRTTPTTAAKMKETATTTTEKTTESKITTATTOY 917
DB      901 tkptaatkpemttakdkterdlrttptttaakmketattttektteskittattoy 960
QY      918 TSTTQDTTPEKITTLLKTTTTLAPKVTTKKITTTEINMKPEETAKPKDRATNSKATPK 977
DB      961 tsttqdttpkittllktttllapkvttckkitttealmkpeetakpkdratnskatpk 1020
QY      978 POKPTAAKPKPSTKPKPTMPVRKKTTPPKKTTSTMPELNPTSRJAEMLOTTTRPN 1037
DB      1021 pokptaaekpkpstpkpmpvrkkttpkkttstmpelnptsrjaemlootttrpn 1080
QY      1038 OTPNKSLVEVNRKSDAGAGETPHMLLRPHVPEVETPMDVTLPRVNGGIIINPMLS 1097
DB      1081 qtpnkslvevnrsdagagetphmlrrphvpevetpmdvtlprvnggiiinpmls 1140
QY      1098 DETNINCKRPVDTGLTLNGLTVAPRGHYFWMLSPFSPSPARRITEVWGISPIDVFT 1157
DB      1141 detnincgkpvdtgltlnglvtvaprghyfwmlspfspsparritevwgispidvft 1200
QY      1158 RCNCEGKFFFKDQOYVFTNDIDAGPRPIFGDGLGOIYAAIASTAKKYNWPESVY 1217
DB      1201 rcncegkffffkdqyvrftndidaagypkpiifgglgylvaalstakynwpeasy 1260

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:25:13 ; Search time 138.84 seconds
(without alignments)
726.115 Million cell updates/sec

Title: AA7
Perfect score: 7323
Sequence: 1 MAWKLPYLLLSVFIQ.....ARAITRSQGLSKWVWNC 1361

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SID58/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID58/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID58/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID58/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID58/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID58/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID58/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID58/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID58/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID58/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID58/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID58/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID58/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID58/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID58/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID58/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID58/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID58/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID58/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID58/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID58/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID58/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7291.5	99.6	1404	AA26049	MSF precursor. S
2	7291.5	99.6	1404	AA26056	Human megakaryocyt
3	7291.5	99.6	1404	AA26057	Human megakaryocyt
4	6715.5	91.7	1299	AA26322	Human EST encoded
5	3484	47.6	902	AA26378	Human MSF-derived
6	2228.5	30.4	452	AA26041	Human megakaryocyt
7	1707.5	23.3	472	AA26056	Bovine MSF ortholo
8	1194	16.3	5179	AA26516	C899P predicted am
9	950	13.0	763	AA26942	Arbidopsin thalia
10	794.5	10.8	1325	AA26306	C. thermophilum OI
11					Peptide #2327 enco

12	756	10.3	182	AA10872	Protein encoded by
13	715	9.8	763	AA26185	Mycobacterium tube
14	643	8.8	4412	AA26366	Sequence g1/01742
15	613	8.4	111	AA26050	MSF-K130. Synthet
16	612	8.4	572	AA26185	Mycobacterium tube
17	552.5	7.5	844	AA26057	Sequence of the fa
18	542	7.4	807	AA26467	Amino acid sequenc
19	520.5	7.1	788	AA26466	Amino acid sequenc
20	510.5	7.0	1837	AA261726	Cryptosporidium pa
21	506.5	6.9	744	AA262975	Bioadhesive precur
22	496.5	6.8	1721	AA268299	Cryptosporidium pa
23	496.5	6.8	1721	AA261727	portion of Cryptos
24	486.5	6.7	2972	AA261231	Human ORFX ORF935
25	489.5	6.7	3118	AA260362	Human SRCAP. Homo
26	488.5	6.7	826	AA26042	Human SRCAP. Homo
27	488.5	6.7	617	AA261458	P. yoelii SSP2 ant
28	488	6.7	617	AA261487	Peptide #2869 enco
29	488	6.7	957	AA269288	Human MUC11 polype
30	488	6.7	957	AA264513	C900P predicted am
31	488	6.7	957	AA264513	Human protein sequ
32	485	6.6	1127	AA269541	Bioadhesive precur
33	476.5	6.5	652	AA262974	Peptide #1317 enco
34	467.5	6.4	511	AA264863	Peptide #1319 enco
35	467.5	6.4	511	AA267312	Peptide #1289 enco
36	467.5	6.4	511	AA262607	Human 07CG27 gene
37	451	6.2	2819	AA263408	PRP 378. Trificum
38	450.5	6.2	378	AA264160	Peptide #967 enco
39	449.5	6.1	2665	AA269550	Peptide #987 enco
40	449.5	6.1	2665	AA262359	Peptide #941 enco
41	449.5	6.1	3265	AA264291	Human ORFX ORF2255
42	446.5	6.1	378	AA264162	PRP encoded by clo
43	446.5	6.1	750	AA265477	C. albicans Rbt1 p
44	446.5	6.1	750	AA265477	Human atrophin-1 r
45	445.5	6.1	1012	AA2617406	

ALIGNMENTS

AA26049	1	AA26049	standard; Protein; 1404 AA.
AA26049		AA26049	
02-FEB-1993	(first entry)		
MSF precursor.			
Megakaryocyte colony stimulating factor; secretion signal; meg-CSF;			
stability; proteolytic cleavage; adhesion; alternative splicing.			
Synthetic.			
Key	Location/Qualifiers		
Region	1..26	/label= Exon_I	
Region	26..67	/label= Exon_II	
Region	67..107	/label= Exon_III	
Region	107..157	/label= Exon_IV	
Region	157..200	/label= Exon_V	
Region	200..1141	/label= Exon_VI	
Region	1411..1166	/label= Exon_VII	
Region	1166..1212	/label= Exon_VIII	
Region	1213..1266	/label= Exon_IX	
Region			

Protein encoded by
Mycobacterium tube
Sequence g1/01742
MSF-K130. Synthet
Mycobacterium tube
Sequence of the fa
Amino acid sequenc
Amino acid sequenc
Cryptosporidium pa
Bioadhesive precur
Cryptosporidium pa
portion of Cryptos
Human ORFX ORF935
Human SRCAP. Homo
P. yoelii SSP2 ant
Peptide #2869 enco
Human MUC11 polype
C900P predicted am
Human protein sequ
Bioadhesive precur
Peptide #1317 enco
Peptide #1319 enco
Peptide #1289 enco
Human 07CG27 gene
PRP 378. Trificum
Peptide #967 enco
Peptide #987 enco
Peptide #941 enco
Human ORFX ORF2255
PRP encoded by clo
C. albicans Rbt1 p
Human atrophin-1 r

p. 1-3

XX WPI: 1986-094065/14.
 DR N-PSDB: AAN60473.
 XX
 XX DNA coding for Plasmodium falciparum antigens - expressing
 PT poly:peptide(s) having antigenicity of RESA or FIRA antigens of P
 PT falciparum
 XX
 PS Disclosure: Fig 7: 55pp: English.
 XX
 CC The inventors claim a novel DNA molecule which comprises a
 CC nucleotide sequence corresp. to all or a portion of the base
 CC sequence coding RESA (AAN60472) or FIRA (AAN60473). RESA and FIRA have
 CC antigenicity suitable for providing protective immunity against
 CC Plasmodium falciparum malarial infections.
 XX
 SQ Sequence 844 AA:

Query Match 7.88; Score 547.5; DB 7; Length 844;
 Best Local Similarity 26.68; Pred. No. 6.5e-24;
 Matches 248; Conservative 118; Mismatches 359; Indels 209; Gaps 49;

```

QY 99 DYSCAFVKNKKKRTKPKRPVVDGSGLDGDFKVTTPDSTQHNKVSPPK 158
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 16 EYSSLSNENPQNBATLTLTPDQ-LSALNHLPN-ELNLETSLTLNENENP 72
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 159 ITTAKPINRPSLPNS-----DTSKETSIVNKETVETKETTNTKOTSTOCKE 209
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 73 VPSAIS-NTLDTLCPYQIIIIISIVNFCRKSSQTYNK--NFEKFNLASVGSNATQGE 129
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 210 KTTSAKETQSIETKISANDLAPTSVLAKPTPKAETTTGPAITTPKEPT---PTTPKEP 265
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 130 NANGNKELNEV-KESSGTQPV-----TPGETVLTG--AAATGETVETGEVTLLEP 179
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 266 ASTPKPEPTTTISAPTPKEPAITTTKSAPTPKEPAITTPKEPAITTPKEPA-----P 321
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 180 VTL-GEVTLT---GEPLTQEP--VLTGEPVLTQGE--PVTQGE--PVTQGEVTLTQEP 227
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 322 TTTKEPAITTTKSAPTPKEPAITTPPKKPAITTPKEPAITTPKEPAITTPKEPAITTPKEP 381
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 228 VLTGEP--VLTGEPVLTQGE-----PVTQGE--PVTQGE--PVTQGEVTLTQEP 269
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 382 A-----PTTPKEPAITTPKEPAITTPKEPAITTPKEPA-----SPTTPKEP 429
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 270 VLTGEPVTLLEPVTL---GEVTLLEP--PVTQGEVTLLEPVTGEPVTLTQEP 323
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 430 APTTTKSAPITTPKEPAITTP-----TTKSAPITTPKEPSPTTPKEPA----- 468
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 324 ---VLTGEPVTLTQGEVTLLEHIDEKKSGEGDNISLSISEETEEKSHLKKKSSWLKFGYR 380
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 469 -----PTTPKEPAITTPPKKPAITTPKEPAITTPPKKPAITTPPKKPAITTPKEPAITTPK 521
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 381 GNKNDKSKNKKPSLESVKNADQEGKPLDSQISVNAQDSVLTGEPALTQEP-PLTGE 439
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 522 ETAVTTPKKTLPITTPKEPAITTPKEPAITTPPEELAPITTPPEEPITTPPEEPAPITTPKAAP 581
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 440 LTA-----TGE--PLTGE--TVEGEPLTGE-----TVAQEP 470
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 582 NTPKEPA-----PTTPKE-----PAITTPKEPAITTPKETAITTPKGTAPITLKEBAPITPK 633
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 471 Iltqepvtlqepvtlqelatlqepstlqehadekkasegdnlsrlseeteekshlkkk 530
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 634 KPAKELAPITTPKEPTSTSDKRAPITTPKGTAPITTPKEPAITTPKEPAITTPKGTAPITL 693
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 531 KSWLKGYNKNDKSKNKKPSLESVKNADQEGKPLDSQISVNAQDSVLTGEPALT 590
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 694 KEBAPITTPKAPAKELAPITTPKGTSTSDKRAPITTPKETAITTPKEPAITTPKAPITTPK 753
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 591 qep-pltqelatlqep--pltq-----etvleqeppltqel--vtlqeppltc---qepvtl 637
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 754 PETPPPTTSVSTPTTTPKEPTTHKSPDE-----STPELSAPETTPKA----- 795
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

```

Db 638 aq-epvtlqel---iatqeppltqehadekkasegdnlsrlseeteekshlkkkssw 693
QY 796 -----LENSPKPEPGVPTTTPKATPKPEMTTAKDXTTERDLTTPPE-TTTAPK 843
Db 694 lKfgynkndkksknekkpslesvknadqegkpldsqisvnaqdsvtlqepaltqep 753
QY 844 MTKETATITTEKTESKITATTTQVTSITTYQDTTPPK--ITTLKTTTTLAPKVTT-----TK 896
Db 754 tlqelatlqeppltqep---tlveqeppltqelvtlqeppltqepvtlqelatlq 810
QY 897 KTTTTIEMKPEETAPKPDRAATNSKATTPKPK 930
Db 811 epsttqge--hadekasegdnlsrlseeteek 842

```

Search completed: April 26, 2002, 16:25:54
 Job time: 544 sec


```

Db      489  appapmpsaavrvpppprlppapppaaprasmpalppapppaptrllcpplppspapnsp 548
Oy      747  KKEAPRTPTPTPTPTSEVSTPTTKEPTTIHKSP-----DESTEPESAEPTPKA--LENSP 800
Db      549  --pppppplppklls--anppcpvppapnrvppapppapppalppapppdpprvansp 604
Oy      801  KKPVPVPTTKTPAATKPEMTTAKDKTTEEDLTETPTTAAKMKMKERATTTTEKTESKI 860
Db      605  --pappapppapppalpfvnpa-----ppcpaapK-----srpal 637
Oy      861  TATTQVSTTQDTTPPKITTLTKTTTLAPKVTTTKTITTEIMNKPEETAKPKDRATN 920
Db      638  paappapppavratcp-----ppapppapppn 665
Oy      921  SKATTPKOKP-----TKAPKKPTSKKKRKTMPRVAKPKTTTPPKMTSTM-----PELNP 971
Db      666  smalpappppdplllatcpappppplpmspppplppaapppppllnqpspplapp 725
Oy      972  TSRIAEAMLQTTTRPNQTPNSKLVENPKSEDAAGAEGET 1011
Db      726  vppaplaplplngprvfarknsl-----gsssgdt 756

```

RESULT 13

AAV53666 standard; Protein: 4412 AA.

AAV53666;

22-FEB-2000 (first entry)

Sequence g1/1017427/emb/CAA62189 from an alignment with protein 608.

Mechanical stress; gene therapy; protein 608; osteoporosis; bone density; bone development; g1/1017427/emb/CAA62189.

Unidentified.

W09960164-AA1.

25-NOV-1999.

14-MAY-1999; 99MO-US11066.

15-MAY-1998; 98US-0085673.

(OUAR-) QUARK BIOTECH INC.

Elmat P, Mor O, Skalliter R, Feinstein E, Faerman A:

WPI; 2000-053304/04.

Identification of stress induced genes for determining risk and preventing, treating or controlling osteoporosis

Claim 32; Fig 6A-R; 308pp; English.

The present sequence is obtained from a clustal X alignment with protein 608. Protein 608 was identified using the method of the invention after subjecting rat osteoblasts to mechanical stress. Expression of the 608 gene was found to be upregulated by about 3-fold in cells subjected to mechanical strain. The specification describes a method for the identification of genes responsive to a specific mechanical stress. The method comprises applying the mechanical stress to an organism (tissue or cells comprising bone cells), isolating the differentially analysing the fractions and extracting mRNA from them, and differentially analysing the mRNA in comparison with control samples. The method is used to identify genes whose expression is responsive to a specific stress. The identified genes are employed in determining risk associated with a physiological or disease state. The risk determination methods are used for testing a medicament for gene therapy. These medicaments, or genes identified by the method of the invention, are used for treating, preventing or

controlling a physiological or disease state (especially osteoporosis or bone density) or other factors causing or contributing to osteoporosis or its symptoms or other conditions involved in mechanical stress or its lack. The methods can also be used for advancing research or studies in bone development.

Sequence 4412 AA;

Query Match 8 9%; Score 627; DB 21; Length 4412;

Best Local Similarity 24.1%; Pred. No. 1e-27; Matches 266; Conservative 110; Mismatches 442; Indels 288; Gaps 51;

```

Oy      67  ELSCGRCFES-----FERGRCDCDAQCKKVDCCPDYESFCA--EYKDNK----- 112
Db      3278  qlschvgseplrlqwlkage-----lkpsdcstsfasgtavlelrdakdsdyv 3331
Oy      113  -----NTRKKKPPK--PPVDEA--GSGLDNGDFKVTTPPTSTTHNKKVSTSPKIT 160
Db      3332  cksavagsdtkakvldkdpavapakkaavdgrlffvsepsqlrvvektctafiakv 3391
Oy      161  TAKPDPNRPSSLPPNSDTSKETSLEVNR-----ETVETKETTNN----- 200
Db      3392  ggdpl-----puvktkqkvrlnggvrflngkddeaklelrdtcltdsglyrcvat 3444
Oy      201  -----KOTSDGKEKTTSAKETOSIETSA-----KDLAPT 231
Db      3445  nehgelesnvnlgvderkkqekleg-dlramlkkllpkkkgageeeldimellknvdpx 3503
Oy      232  S-----KYLAPTPKAETTTKGA 250
Db      3504  eyekyarmyldtfrglqgvveehvkvrvleafeaevefepk-----kappgpe 3559
Oy      251  LT-----TPKEPTPTTP--KEPASTTPKEPTTIKSAPTTPKEP--APTTPKSAPTTPKEP 303
Db      3560  lseklipkkpckkvprkvpkvpkvpkvpkvpkvpkvpkvpkvpkvpkvpkvpkvpk 3619
Oy      304  APTTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPA 359
Db      3620  VP-----ekkvppakkeap-----ppkvpeapkeavpkeavpvpkvpkvpkvp 3665
Oy      360  PTPKEPTPTTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPA 412
Db      3666  tkvpevpkaavpkeavpkeavpkeavpkeavpkeavpkeavpkeavpkeavpkeav 3724
Oy      413  PAPTTPKESPTTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPA 472
Db      3725  vvpkvpvpaap--pkkevprvk--vpeapkeavpkeavpvp--ppkkvprvk-----vp 3775
Oy      473  KKPAPTPPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPK 532
Db      3776  evpkvavpkeavp-----ealppkespppvefeeevvalieppaaveeppaappqv 3831
Oy      533  ----PTTPKEKLAPT--PEKPAPTPPELAPTTPEEPTP--TPPEAPTPPKAAPTNT 583
Db      3832  vppknpvpekkapavavakpelpvkvpevpkeavpkeavpkeavpkeavpkeavpke 3889
Oy      584  PKEPAP-----TPKEP-----APTTPKPAATTPKPAATTPKPAATTPKPAATTPK 633
Db      3890  pkeavpkeavpkeavpkeavpkeavpkeavpkeavpkeavpkeavpkeavpkeavp 3942
Oy      634  KPAPELAPTTTKEPTSTSDKPAATTPKGTAPTTKPAATTPKPAATTPKPAATTPKPA 693
Db      3943  e--peelap-----eeelapeekpvyvae-----eeevpppavpeekllpckvvp -v1 3992
Oy      694  KKPAPTPPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPK 746
Db      3993  kkppepppkepekev-----lekplkprpppppppdkvexklfqlkalpkkvpevp 4049
Oy      747  KKP--APTTPPTPTTSPVS--TPTTPKEP-----TTIKSDESTPELSAEPTRKA 795
Db      4050  qvpekveltpkvpggkvrkllperlpkekeavvllksvllrkpreeeepkve-----pkk 4105

```



```

QY 408 TPKKEPA--PTTKESPT--TPKEPAPTTKS--APTTKEPAPT---TTKSAPTTPKES 460
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 920 sdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdepl 979
QY 461 ----PTTKEPAPT--TPKEBAPTTPKKAPPTTPKEBAPTTPKEBAPTTPKKAPPTAPKBP 515
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 980 psdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdepl 1039
QY 516 APT---TPKEBAPTTPKKAPPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPT 568
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1040 tpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdepl 1094
QY 569 EEPAPTTPKAAPNTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPT 625
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1095 eep1p1psdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdepl 1154
QY 626 EEPAPTTPKKAPKELAPTTTKAPT---STSDKAPPTTPKGTAPPTTPKEBAPTTPKEBAPT 684
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1155 eptps--deplpsd--eptpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdepl 1208
QY 685 PKGTAPTTKEBAPT--TPKKAPKELAPTTTKGPTSTSDKAPAPT---TPKETAAPTTPKE 740
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1209 ----deplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsde 1253
QY 741 PAPTPPKKAPPT--TPETPPTTSEVSTPTTKAPTTPKSPDESPTPELSAPPTPKALENS 799
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1254 ptps--deplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsde 1305
QY 800 PKEPVPPTTKTPAPT--KPEMTTAKDKTTERDLATTPETTTAANKMKETATTEKTE 857
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1306 sdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdepl 1361
QY 858 SKITTTTQVSTTTQDTTPPKITTTLLKTTTLAKVTTTKKITTTEIMNKPEETAKPKDR 917
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1362 deplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdepl 1396
QY 918 ATNSKAT-----TPKQKPTKAP--KKPSTKKPKMKPVKPKKPTTPTRKM----- 962
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1397 sggsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsg 1453
QY 963 -----TSMPELNPSTSRILAEMLQTTTPRNPQTPNSKLVENPKSEDAGAEGETPHM- 1014
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1454 yllgydgsfipernittraaavf-----akll-----gadesyagsgaspsd 1498
QY 1015 LLRPH-----VMPR-----VTPMDYLRPRVNGI----- 1040
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1499 lachmaawaikfatsgqlfkgydgtfkpqnlttraefatvhlfltkvvgelmskla 1558
QY 1041 ---IINPULSDETNICNG---KPVDTLTL-----RNGTLVAFRGHIFMMLSPSPSPS 1087
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1559 tldisnpkfd---cvgnhaqefiekltsgylsgypdgt-----fkpqn 1600
QY 1088 PARITEWAGPSPIYVTRFCNCEGKTFFPKD--SQYWRFTINDIKD 1132
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1601 ylkrsav---allnralerplngapklfpdvnesyaf-gdlmd 1642

```

```

RESULT 11
AA03645
ID AA03645 standard; Protein: 1325 AA.
XX AA03645;
XX
XX 09-OCT-2001 (first entry)
XX
XX Peptide #2377 encoded by probe for measuring breast gene expression.
XX
XX Probe: human; breast disease; breast cancer; development disorder;
XX Inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
XX Homo sapiens.
XX
XX MO200157270-A2.

```

```

XX
PD 09-AUG-2001.
XX
PF 29-JAN-2001; 2001WO-US00661.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI: 2001-476286/51.
XX
XX Novel single exon nucleic acid probe used to measuring gene expression
XX In a human breast -
XX
XX Claim 27; SEQ ID NO 12385; 322pp; English.
XX
XX The present invention relates to novel single exon nucleic acid probes
XX (see A100010-A110067). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for measuring human gene expression in
XX a human breast sample, where the probe hybridises at high stringency to a
XX nucleic acid expressed in the human breast. The probes are useful for
XX predicting, diagnosing, grading, staging, monitoring and prognosis
XX diseases of the human breast, particularly those diseases with polygenic
XX aetiology. The diseases include: breast cancer, disorders of development,
XX inflammatory diseases of the breast, fibrocystic changes, proliferative
XX breast disease and non-carcinoma tumours.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1325 AA:

```

```

Query Match 11.0%; Score 776.5; DB 22; Length 1325;
Best Local Similarity 25.9%; Pred. No. 5.8e-37;
Matches 304; Conservative 99; Mismatches 492; Indels 281; Gaps 43;
QY 76 ESPEREGECDCQCKKYKCCPDYSCAEVKNKNTKKKPRPKPVVDAGSGLDN 135
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 51 elheqkthcrlthskprdk--ptgnsktldhksctdn-----eapltseensngk 102
QY 136 GDF-----KVTPDSTTOHNVKSTSPKITTA--KPINRPSLPRNSDYSK-----E 180
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 103 dpmllngqsvdpsdstltthkesgkhhlpapbsklnctsktsgkvtlrvkadtgrple 162
QY 181 TSLTVKKEKTEVTEKETT-----NKQSTNGKEKTTSAKET-----QSIETK- 223
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 163 kmsltdktseshklttsfhngsqtkqsktsfpekltlaaskltkyktgtpseesekte 222
QY 224 SAKDLAPTISKVLAKPRPKKETTTPKGPALN-----TPKEPPT----- 260
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 223 dsrttvaasklltktknigtlsanelqtslaepelhygrtanennnpspaeptenrer 282
QY 261 TPKEPASTTPKEPPTTTTISA--PTTPKEBAPT-----TTSAPPTTPKEBAPTTPK 309
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 283 tanentlpsaeptenrerlanentapfapgltenemtanentlcfpaeplehertan 342
QY 310 E---PAPTPPKAPPTTPKEBAPTTPKSAPTTPKEBAPTTPPKAPPTTPKEBAPTTPKEP 366
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 343 entcpsaeplehertane-----nltpsaepelhygrtlfandktsssestehegr 398
QY 367 TP-----TPKEBAPT--TPKEBAPTTPKEBAPTAPKAPPTTPKEBAPTTPKEBAPT--PTT 417
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 399 tplanentpsaeptenrerlanentcpspaeptenretlanektlispeptenret 458

```

PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 26-OCT-1999; 99US-0161362.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161921.
 PR 28-OCT-1999; 99US-0161922.
 PR 28-OCT-1999; 99US-0161923.
 PR 29-OCT-1999; 99US-0162142.

Query Match 13.98; Score 981; DB 21; Length 763;
 Best Local Similarity 38.58; Pred. No. 4.4e-49;
 Matches 226; Conservative 24; Mismatches 293; Indels 44; Gaps 11;

QY 243 ETTTKEGALTTPKEPTTTPKEPASTTPKEPTTITKSAPTTPKEPASTTPKE 302
 DB 126 qtdagurpylirppprirpppcvclcp-esp------pppvtl---pq1ppt 170
 QY 303 PAPTTPKEPAPTTPKEPAPTTPKEPAPT--TKSAPTTPKEPAPTTPKEPAPTTPKEPAP 360
 DB 171 lrpptlppetppppapppcevlptpprlppqtpcllppelppppqtpcllp---p 226
 QY 361 TTPKEPTTPKEPAPT--KEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPK 419
 DB 227 ntpetppcllppetppcllppetppcllppetppcllppetppcllppetppcllpp 286
 QY 420 EPSPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKEPAPTTPKEPAPT 479
 DB 287 etupntppescllppelppntppsp-----ntppetppntppptppntppcllppnt 340
 QY 480 PKKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKE 539
 DB 341 ppgspntppetppcllppetppcllppetppcllppetppcllppetppcllppet 400
 QY 540 APTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 599
 DB 401 lppntppcllppetppcllppetppcllppetppcllppetppcllppetppcllpp 460
 QY 600 PREPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKE 659
 DB 461 pklp-ppcllppetppcllppetppcllppetppcllppetppcllppetppcllpp 514
 QY 660 TPKGTAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKE 719
 DB 515 tppgtppnqspqtpcllppntppetppcllppntppetppcllppntppetppcllpp 574
 QY 720 TTSDEP-----APTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKE 772
 DB 575 lppntppcllppntppcllppntppcllppntppcllppntppcllppntppcllpp 634
 QY 773 P---TTIHKSDESTPELSAEPPTKALINSKEGCVPTTKKPAATKP 816
 DB 635 pppttllppspptppnpspppksqpppprpfqpppppgrlcp 681

RESULT 10
 AAM43106
 ID AAM43106 standard; Protein; 1664 AA.
 AC AAM43106;
 XX 16-OCT-1998 (first entry)
 DT
 XX
 DE C. thermocellum OLPB protein.
 XX
 XX Multimer; enzyme; complex; protein-protein interaction; dockerin domain;
 KM cohesin domain; catalytic subunit; scaffold subunit; SdbA; synergistic;
 XX cellulosome integrating protein; scaffoldin dockerin binding protein.
 OS Clostridium thermocellum.
 XX
 XX
 FH Key Location/Qualifiers

FT Domain
 FT /note="cohesin type II domain"
 FT 207..363
 FT Domain
 FT /note="cohesin type II domain"
 FT 409..565
 FT Domain
 FT /note="cohesin type II domain"
 FT 607..763
 FT Domain
 FT /note="cohesin type II domain"
 XX
 XX FR2748479-A1.
 XX
 XX
 PD 14-NOV-1997.
 XX
 XX
 PR 10-MAY-1996; 96FR-0005854.
 XX
 XX 10-MAY-1996; 96FR-0005854.
 XX
 XX (INSP) INST PASTEUR.
 XX
 PI Beguin P, Leibovitz E;
 XX
 DR WPI; 1998-011569/02.
 XX N-PSDB; AAT86623.
 XX

Cellulase proteins with cohesin or dockerin type II domains - useful for potentiating the activity of multiprotein enzyme complexes
 Claim 7; Page 31-39; 60pp; French.

CC Multimeric protein, especially enzymatic, complexes are held together
 CC by protein-protein interactions between domains designated dockerins
 CC and cohesins, which are found on the catalytic and scaffold subunits
 CC respectively. An example of such a complex is the cellulose degrading
 CC protein complex from Clostridium thermocellum, known as the cellulosome.
 CC This complex comprises around 15 proteins including endoglucanases,
 CC cellobiohydrolases, hemicellulases, e.g. xylnases or lichenases, which
 CC interact with a central "scaffold" protein designated the cellulosome
 CC integrating protein (CipA; see AAM43108). The catalytic subunits
 CC interact with the Cip subunit via conserved 23 amino acid dockerin
 CC domains. Cip has been shown to contain 9 copies of a cohesin domain.
 CC The invention relates to the isolation of proteins binding to a novel
 CC dockerin type domain found in the C-terminal portion of Cip. The new
 CC I domain found on the catalytic subunits of the cellulosome) The type
 CC II dockerin domain has some sequence similarity to the type I dockerins
 CC but is unable to bind type I cohesin domains.
 CC The sequence presented here is an example of a protein which binds
 CC the novel type II dockerin domain and is the product of the OLPB gene.
 CC The protein contains 4 type II cohesin domains in the N-terminal portion
 CC of which the first domain (amino acid residues 28-192) is thought to
 CC bind CipA. The novel type II dockerin and cohesin domains can be used
 CC in complexes, especially enzyme complexes, to potentiate their catalytic
 CC actions in a synergistic manner.
 XX
 XX Sequence 1664 AA:
 SQ

Query Match 13.48; Score 950; DB 19; Length 1664;
 Best Local Similarity 31.08; Pred. No. 6.6e-47;
 Matches 312; Conservative 104; Mismatches 361; Indels 230; Gaps 55;

QY 234 VLAQTP-KAETTTKGPALTTPKEPTTPKEPASTTPKEPTTPITKSAPTTPKEPAPTTP 292
 DB 758 vwidgapikaasdepilpdlpsdepts-----depts-----deptsdeptsd 804
 QY 293 TKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPK 348
 DB 805 epupsetceepilpdlpsdeptsdeptsdeptsdeptsdeptsdeptsdeptsdepts 864
 QY 349 KPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKE 407
 DB 865 sdeptsdepts--depts--depts--deptsdeptsdeptsdeptsdeptsdepts 919

PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139464.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140655.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140951.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0143542.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150566.
PR 27-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154779.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-015659.
PR 29-SEP-1999; 99US-0156458.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158025.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.


```

XX DE Bovine MSF orthologue, superficial zone protein (SZP).
XX XX
XX KW Bovine; CACP protein; campodactylly-arthropathy-coxa vara-pericarditis;
XX KW superficial zone protein; SZP; MSF; superficial zone protein;
XX KW osteoarthritis; joint lubrication; osteopathic; ankiarthritic.
XX OS Bos taurus.
XX PN WO200107068-A1.
XX PD 01-FEB-2001.
XX XX
XX PE 21-JUL-2000; 2000WO-US20002.
XX XX
XX PR 23-JUL-1999; 99US-0145328.
XX PR 19-JUL-2000; 2000US-0145328.
XX PA (UYCA-) UNIV CASE WESTERN RESERVE.
XX PI Warman ML.
XX DR WPI; 2001-182721/18.
XX PT New composition comprising the campodactylly-arthropathy-coxa
XX PT vara-pericarditis protein in combination with an anesthetic, useful for
XX PT treating osteoarthritis, or as lubricants of tissue and joints
XX XX
XX PS Example 1; Fig 4; 34pp; English.
XX XX
XX CC The invention relates to a method of treating osteoarthritis via the
XX CC administration of a composition comprising the campodactylly-arthropathy-
XX CC coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.
XX CC The composition may further comprise a local anesthetic. The composition
XX CC of the invention may be administered via intra-articular or intravenous
XX CC injection. The human CACP protein is identified in the invention as
XX CC being megakaryocyte stimulating factor (MSF). The gene encoding
XX CC CACP protein (MSF) is located on chromosome 1q25-31, and mutations in
XX CC this gene are responsible for the heritable disorder campodactylly-
XX CC arthropathy-coxa vara-pericarditis, in which patients have synovial
XX CC hyperplasia without evidence of inflammation. CACP protein (MSF)
XX CC acts as a synovium lubricant, and can be used to lubricate tissue and
XX CC joints in the treatment of osteoarthritis. The composition may be
XX CC applied to reduce the symptoms of osteoarthritis (e.g., joint pain,
XX CC loss of range of movement or joint damage). The present sequence
XX CC represents the bovine orthologue of human MSF, superficial zone
XX CC protein (SZP).
XX XX
XX SQ Sequence 472 AA;

```

```

Db 314 alsiaqksrpesvfyfkrsgsvqgytykqepbctctgrrpalnysvgetaqvrrrrfe 373
Qy 1213 RAIGSQHTTIRIOYSPARLAYQDKGVLANEKKVSLMRGLPNVWTSATSPNIRKPPDGY 1272
Db 374 raigsqvhtlrlhytprvpygdkgflhnekvslwrglpnvvtaslsipnlrkpogy 433
Qy 1273 DYAFSKDOYNNIDVPSRTARAITTRSGOTLSKWYNCP 1311
Db 434 dyafskdgyynidvpsrtaraittrsgotlsnwnyep 472

RESULT 8
AAM24516
ID AAM24516 standard; Protein; 5179 AA.
XX XX
XX AC AAM24516;
XX XX
XX DT 12-OCT-2001 (first entry)
XX XX
XX DE C899P predicted amino acid sequence.
XX XX
XX KW Human; immunotherapy; diagnosis; colon cancer; colon tumour;
XX KW immunogenic; gene therapy; vaccine; colonic cancer.
XX XX
XX PA Homo sapiens.
XX PN WO200149716-A2.
XX PD 12-JUL-2001.
XX XX
XX PE 29-DEC-2000; 2000WO-US35596.
XX XX
XX PR 30-DEC-1999; 99US-0476296.
XX PR 10-JAN-2000; 2000US-0480321.
XX PR 15-FEB-2000; 2000US-0504629.
XX PR 06-MAR-2000; 2000US-0519444.
XX PR 19-MAY-2000; 2000US-0575251.
XX PR 29-JUN-2000; 2000US-0609448.
XX PR 28-AUG-2000; 2000US-0649811.
XX XX
XX PA (CORI-) CORIXA CORP.
XX XX
XX PI Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
XX PI King GE, Wang T, Jiang Y;
XX XX
XX DR WPI; 2001-441847/47.
XX XX
XX PT Colon tumor associated proteins and nucleic acids useful for the
XX PT prevention, diagnosis and treatment of colonic cancer -
XX PS Claim 2; Page 446-462; 472pp; English.
XX XX
XX CC The present invention describes colon tumour associated proteins (I) and
XX CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
XX CC (I) and (II) may be used in gene therapy and vaccine production. (I) and
XX CC (II) may be used in the prevention, diagnosis and treatment of diseases
XX CC associated with inappropriate colon tumour associated protein (TCAP)
XX CC expression, such as colonic cancer. For example, (I) and (II) may be
XX CC used to treat disorders associated with decreased expression by
XX CC rectifying mutations or deletions in a patient's genome that affect the
XX CC activity of TCAPs by expressing inactive proteins or to supplement the
XX CC patients own production of them. Additionally, (II) may be used to
XX CC produce the TCAP proteins, by inserting the nucleic acids into a host
XX CC cell culturing the cell to express the protein. (II) and its
XX CC complementary sequences may also be used as DNA probes in diagnostic
XX CC polymerase chain reaction (PCR) and hybridisation assays to detect and
XX CC quantitate the presence of similar nucleic acids in samples, and
XX CC therefore which patients may be in need of restorative therapy. (I) may
XX CC also be used as antigens in the production of antibodies against TCAPs
XX CC and in assays to identify modulators of TCAP expression and activity.
XX CC Anti-(I) antibodies and antagonists may also be used to down regulate
XX CC TCAP expression and activity. The anti-(I) antibodies may also be used
XX CC as diagnostic agents for detecting the presence of TCAPs in samples

```



```

Db      721 aaptpkpkpapktapcttkeptsctsdckpaptckgtaptckpaptckpaptckpaptckp 780
OY      688 TAATTLKEAPATPPKPAKELAPTTKGTSTSDKPAATPPKAPATPPKAPATPPK 747
Db      781 taatlckepatkpkpapelapcttkgpcstsdckpaptckpaptckpaptckpaptckp 840
OY      748 KPAATPEPTTSEVSTPTTKEPTTIHKSPESTPELSAEPKALENSPEKGEVPT 807
Db      841 kpaatpceppttsevspttckpaptckpaptckpaptckpaptckpaptckpaptckp 900
OY      808 TKTPTAATPEMTTAKDKTTERDLRTTPEPTTAAPKMTKATATTTKTESKTATTTTQY 867
Db      901 tktptaatkpemtatakdktterdlrttpepttlaapkmtkattctktskitalcttqy 960
OY      868 TSTTTODTPEFKITTLTKTTTLAPKAVTTTKTITTTTEIMNKEPTAKPKDRATSKATTPK 927
Db      961 tstitqdtcpkiklcltktlapvttktklitttelimkpeetapkdatatnkattpk 1020
OY      928 POKPTKAPKPTSTTKKPKTMPRVKRPKPTTPPKMTSTMPDELNPTSRIAEAMLTQTTTRPN 987
Db      1021 poptkpkpstkstkpktmprvrkptktpckmtstmpelnptsrilaeamlqtltrpn 1080
OY      988 QTNSKILVENPKSESDGAGHGETPHMLLRPHVMPPEVTPDMVLYLPVPMOGIIINPMLS 1047
Db      1081 qtpnsklivenpksedegagelphmllrphvfmpevcpdmdylpvpnngiilnplms 1140
OY      1048 DETNICKGKRPVDGLTTLRNCTLVAFRGHYFMMLSPFSPSPARITTEVWGISPIDTIVET 1107
Db      1141 detnicgkkrpvdglttlrnctlvafrghyfmmlspfsparriltevwgispidtvlft 1200
OY      1108 RCNCEGTFEFPKDSQVRFNTNDIKDAGYPKPIKGGEGGLIGQIYALSTAKYKMPESVY 1167
Db      1201 rcncegtrffkfdqgyrftndldagypkpfkfglqgivaalstakympesvy 1260
OY      1168 FFRKGGSIQOYIYKOEPOVOKOPRRPALNYPVGEVMTQVRRRPERAIGSOTHTIRIQY 1227
Db      1261 ffrkgsigqiykqepvqkcparrpalnypvgemqvttrrrferaiqpsqhtlrlqy 1320
OY      1228 SPARLAAYODKGVILHNEKVSILMRGLPNVYVTSATSLPNIRKPDGYDYAASKDOYINIV 1287
Db      1321 sparlalayqdkgvilnehkvslwrglpnvvtlsaislpnirkpdgydyafskdgyndlv 1380
OY      1288 PSRTARAITTRSGOTLSKWTNCP 1311
Db      1381 psrtaraittrsgtllskwvynpc 1404

RESULT 4
AAM24322
ID      AAM24322 standard; Protein: 1299 AA.
XX
AC      AAM24322;
XX
DE      12-Oct-2001 (first entry)
XX
DE      Human EST encoded protein SEQ ID NO: 1847.
XX
KW      Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW      tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW      diagnostics; forensic test; gene mapping; genetic disorder;
KW      biodiversity; gene therapy; nutrition.
XX
OS      Homo sapiens.
XX
PN      WO200154477-A2.
XX
PD      02-AUG-2001.
XX
PE      25-JAN-2001; 2001WO-US02687.
XX
PR      25-JAN-2000; 2000US-0491404.
PR      17-JUL-2000; 2000US-0617746.

```

```

PR      03-AUG-2000; 2000US-0631451.
PR      15-SEP-2000; 2000US-0663870.
XX
PA      (HYSE-) HYSEQ INC.
XX
PI      Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI      Cao Y, Drmanac RA, Zhang J, Werhman T;
XX
DR      WPI, 2001-476164/51.
DR      N-PSDB; AAH98981.
XX
PT      Isolated polypeptide for treatment of diseases, diagnostics, raising
PT      antibodies and research use -
XX
PS      Claim 20; Page 1198-1201; 1275pp; English.
XX
CC      The present invention provides the protein and coding sequences of novel
CC      proteins from a variety of organisms, including human, dog, cat, horse,
CC      cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC      urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC      from the organism of interest. They can be used in diagnostics,
CC      forensics, gene mapping, identification of mutations, to assess
CC      biodiversity and for nutritional purposes. The present sequence is a
CC      protein of the invention.
XX
SO      Sequence 1299 AA;

```

```

Query Match 91.0%; Score 6431.5; DB 22; Length 1299;
Best Local Similarity 92.8%; Pred. No. 0;
Matches 1202; Conservative 0; Mismatches 0; Indels 93; Gaps 1;

```

```

OY      1 MAMKTLPIYLLLLLSVFIQOVSSODLSSCAGCGEGYSRDATCNCNDYMCQHYMECCPDF 60
Db      1 mawklpiyililllsvfiqovssqdlscagcgcgysrdatcncndyhmeccpdf 60
OY      61 KRYCTAELSCGRGCFESFERGERCDCAOCKKYDKCCPDYESFCAR----- 106
Db      61 kryctaelsckgrcfesfergercdcaqckkydkcpdyesfcaevhnptspssakap 120
OY      107 ----- 106
Db      107 ----- 106
OY      121 ppsgaqgltkstkirkpknkkkkkvvieseeleehsvsenqesssssssssstllw 180
Db      107 -----VKDNKKNRKKKKKPPKPVVDAGSGLDNGDFKVTTPDTS 147
OY      181 kixssksaanrelqkklxkdknkutkkkppkpvvdeagsgldngdfkvtlpdst 240
Db      148 TOHNKVTSEKITTAKPINRPSLIPNSDTSKETSLTVNKETTVERKETTITTKKQSTDC 207
Db      241 tqhnkvstspkltakpimpisipnsdstsketsltnkettvetkettltnkqstldg 300
OY      208 KEKTTSAKETOSTIEKTSADLAPTSKYLAKPTPKAETTTGPAITTPKKEPTTPPKEPAS 267
Db      301 kekttsaaketostiektakdlaptskylakptpkaeittgpaaltppkpepttpkpas 360
OY      268 TTPKEPTTITKSAPITPKKEPAPTTPKSAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPK 327
Db      361 ttpkeptitksapitpkkepapttpkkaepttpkkaepttpkkaepttpkkaepttpk 420
OY      328 APTTTSAPTPPKPAATTPPKKPAATTPPKKPAATTPPKKPAATTPPKKPAATTPPKKPA 387
Db      421 aptttsaptpkpaattpkpaattpkpaattpkpaattpkpaattpkpaattpkpaattpk 480
OY      388 EPAPTAPKPAATTPPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPK 447
Db      481 epaptapkpaattpkpaattpkpaattpkpaattpkpaattpkpaattpkpaattpk 540
OY      448 TTKSAPTPKKEPSPTTKEPAPTTPKKEPAPTTPKKAATTPPKKPAATTPPKKPAATTPK 507
Db      541 ttksaptpkkespttkepapttpkkaattpkkaattpkkaattpkkaattpkkaattpk 600
OY      508 APTAKKEPAPTTPKETAATTTKKLPTTPPKLAATTPPKKPAATTPPEELAPTPPEEPTPT 567

```


QY 928 PQRPAKPKPTSTKPKPTMPRVKRPKPTPTPRKMTSTMBELNPTSRIAEAMLOTTTRPN 987
DB 1021 PQRPKAPKPKPTSTKPKPTMPRVKRPKPTPTPRKMTSTMBELNPTSRIAEAMLOTTTRPN 1080
QY 988 QTNNSLVENPKSEDAAGAGETPMMLRPHVMEPVTPDMOYLPRVNOGIIINPMLS 1047
DB 1081 QTNNSLVENPKSEDAAGAGETPMMLRPHVMEPVTPDMOYLPRVNOGIIINPMLS 1140
QY 1048 DETNIGNKRPVDDLTTLRNGTTLVAFRGHYFMMLSPSPSPSPARTRTEWVGISPTDVTFT 1107
DB 1141 DETNIGNKRPVDDLTTLRNGTTLVAFRGHYFMMLSPSPSPSPARTRTEWVGISPTDVTFT 1200
QY 1108 RCMCEKRTFFKDSQYWRNTNDIKAGYKRPJFKGFGSLTGOIYVALSTAKYKMPESYV 1167
DB 1201 RCMCEKRTFFKDSQYWRNTNDIKAGYKRPJFKGFGSLTGOIYVALSTAKYKMPESYV 1260
QY 1168 FFKRGSIQYIKOEPVOKCPGRPALMYPYGEMTOVRRRRRFRRAIGPSTHTIRIOY 1227
DB 1261 FFKRGSIQYIKOEPVOKCPGRPALMYPYGEMTOVRRRRRFRRAIGPSTHTIRIOY 1320
QY 1228 SPARLAYODKGYLVHNEVKYSILMRGLPNVVTSAISLPNIRKPDGYDYAFSKDQYNNIDY 1287
DB 1321 SPARLAYODKGYLVHNEVKYSILMRGLPNVVTSAISLPNIRKPDGYDYAFSKDQYNNIDY 1380
QY 1288 PSRTARATITRSGOTLSKRWYNCP 1311
DB 1381 PSRTARATITRSGOTLSKRWYNCP 1404

RESULT 3
AAB29773
ID AAB29773 standard; Protein: 1404 AA.

AC AAB29773;

DT 28-FEB-2001 (first entry)

DE Human megakaryocyte stimulating factor (MSF), SEQ ID NO:1.

KW Human MSF, megakaryocyte stimulating factor; tribonectin;

KW alternative splicing; joint boundary lubricant; O-linked oligosaccharide;

KW osteoarthritis; triboneupplementation; tissue adhesion inhibition;

KW friction coefficient reduction; gene therapy; anarthritic;

KW osteopathic.

OS Homo sapiens.

PN MO200064930-A2.

PD 02-NOV-2000.

PE 24-APR-2000: 2000MO-US10953.

PR 23-APR-1999: 99US-0298970.

PA (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.

PI Jay GD;

DR WPI: 2001-024673/03.

DR N-PSDB; AAC81498.

PT Novel tribonectin polypeptide useful as lubricant for treating

PS Claim 3; Page 7; 47pp; English.

CC human MSF-derived tribonectin; a biocompatible composition comprising a
CC human tribonectin for inhibiting tissue adhesion formation; and a method
CC of diagnosing osteoarthritis or a predisposition to osteoarthritis by
CC measuring the amount of MSF or its fragment in a biological sample of a
CC mammal, wherein an increased amount of MSF compared to a control
CC indicates the presence of or predisposition to developing
CC osteoarthritis. The tribonectin and DNA encoding it are useful in the
CC treatment of osteoarthritis, where they may be used for lubricating
CC mammalian joints, such as articulating joints of humans, dogs or horses.
CC The tribonectin, when formulated as a membrane, foam, gel or fibre, is
CC useful for inhibiting adhesion between two surfaces such as the injured
CC or trauma, or an artificial device e.g., an orthopaedic implant. In
CC particular, one of the surfaces is pericardial tissue. DNA encoding a
CC tribonectin may be used in gene therapy. The present sequence represents
CC human MSF.
SQ Sequence 1404 AA:

Query Match 99.28; Score 7007.5; DB 22; Length 1404;
Best Local Similarly 93.44; Pred. No. 0;
Matches 1311; Conservative 0; Mismatches 0; Indels 93; Gaps 1;

QY 1 MAWKTLPIYLLLSVFIQOVSSODLSSCAGRGEGYSRDATNCNDYNCQHTWECCPDF 60

DB 1 MAWKTLPIYLLLSVFIQOVSSODLSSCAGRGEGYSRDATNCNDYNCQHTWECCPDF 60

QY 61 KRVCTAELSCGRCFESFERGECDDAQCCKYDKCCPDFESFCAE----- 106

DB 61 KRVCTAELSCGRCFESFERGECDDAQCCKYDKCCPDFESFCAE----- 120

QY 107 ----- 106

DB 121 PPSGASGLIKSTKSPKPKKKKKVLESSEELTEHSVSENSGSSSSSSSSSLW 180

QY 107 -----VKDNKKNRTKKKPKPKPVNDGSSGNDGFKVTTEDTS 147

DB 181 KIKSKNSAANRELQKLLKVKDKKNTKKKPKPKPVNDGSSGNDGFKVTTEDTS 240

QY 148 TOHNKSTSKKITPAKINDPSLPPNSDSKETSITLVNKETTVETKETTNTKOTSTG 207

DB 241 TQHNKSTSKKITPAKINDPSLPPNSDSKETSITLVNKETTVETKETTNTKOTSTG 300

QY 208 KEKTTSAKETOSTEKTSAKDLAPTSTKVLAKPTPKAETTTGCPALTTPKEPTTPKEPAS 267

DB 301 KEKTTSAKETOSTEKTSAKDLAPTSTKVLAKPTPKAETTTGCPALTTPKEPTTPKEPAS 360

QY 268 TTPKEPTTPKTSAPTTPKEPAATTTKSAPTTPKEPAATTTKSAPTTPKEPAATTTK 327

DB 361 TTPKEPTTPKTSAPTTPKEPAATTTKSAPTTPKEPAATTTKSAPTTPKEPAATTTK 420

QY 328 APTTTSAPTTPKEPAATTTKSAPTTPKEPAATTTKSAPTTPKEPAATTTKSAPTTP 387

DB 421 APTTTSAPTTPKEPAATTTKSAPTTPKEPAATTTKSAPTTPKEPAATTTKSAPTTP 480

QY 388 EPAPTAPKPAATTPKEPAATTTKSAPTTPKEPAATTTKSAPTTPKEPAATTTKSAPT 447

DB 481 EPAPTAPKPAATTPKEPAATTTKSAPTTPKEPAATTTKSAPTTPKEPAATTTKSAPT 540

QY 448 TTKSAPTTPKEPSPTTKKEPAATTPKEPAATTTKSAPTTPKEPAATTTKSAPTTP 507

DB 541 TTKSAPTTPKEPSPTTKKEPAATTPKEPAATTTKSAPTTPKEPAATTTKSAPTTP 600

QY 508 APTAPKEPAATTPKEPAATTTKSAPTTPKEPAATTTKSAPTTPKEPAATTTKSAPT 567

DB 601 APTAPKEPAATTPKEPAATTTKSAPTTPKEPAATTTKSAPTTPKEPAATTTKSAPT 660

QY 568 PEPAATTPKAAAPNTKEPAATTTKSAPTTPKEPAATTTKSAPTTPKEPAATTTKSAPT 627

DB 661 PEPAATTPKAAAPNTKEPAATTTKSAPTTPKEPAATTTKSAPTTPKEPAATTTKSAPT 720

QY 628 APTTPKPAKPAATTTKKEPTSTSDKPAATTPKGTAPTTKKEPAATTTKKEPAATTPKG 687

QY	1168	FFKRGSGIQGYITIKOEPPQKCGRRPALNPPYGCMTQVRRRRRPERAIGSGOTHRIRIOY	1227
Db	1261	ffkrsgsldqylykpepqvkcpgrpalnpppygemltvrrrrleraiqpsqthlrrdy	1320
QY	1228	SPARLAYODKGYLHNEVAVSILMRCGLPNVNVSAISLAPNIRKPDGYDYAFSKDOYYNIDV	1287
Db	1321	sparlaysdkgyllhnevksllwrlplnvvsaislplnrkpdgydyafskdyynldv	1380
QY	1288	PSRTARAITRRSGQFLSKWYNCP	1311
Db	1381	psrtaraitrrsgqlskwyncp	1404
RESULT	2		
ID	AAB60568		
XX	AAB60568	standard; Protein; 1404 AA.	
AC	AAB60568;		
DT	27-APR-2001	(first entry)	
XX			
DE		Human megakaryocyte stimulating factor (MSF, CACP).	
XX		Human: CACP protein; campitodactyl-arthropathy-coxa vara-pericarditis;	
KW		MSF: megakaryocyte stimulating factor; synovial lubricant;	
KM		chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;	
KW		antiarthritic.	
OS		Homo sapiens.	
XX			
PN	WO200107068-A1.		
XX			
PD	01-FEB-2001.		
XX			
PF	21-JUL-2000; 2000MO-US20002.		
XX			
PR	23-JUL-1999; 99US-0145328.		
XX	19-JUL-2000; 2000US-0145328.		
PA	(UYCA-) UNIV CASE WESTERN RESERVE.		
PI	Warman ML;		
XX			
DR	WPI; 2001-182721/18.		
XX			
PT	New composition comprising the campitodactyl-arthropathy-coxa		
PT	vara-pericarditis protein in combination with an anesthetic, useful for		
PT	treating osteoarthritis, or as lubricants of tissue and joints		
XX			
PS	Example 1; Page -: 34pp; English.		
XX			
CC	The invention relates to a method of treating osteoarthritis via the		
CC	administration of a composition comprising the campitodactyl-arthropathy-		
CC	coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.		
CC	The composition may further comprise a local anaesthetic. The composition		
CC	of the invention may be administered via intra-articular or intravenous		
CC	injection. The human CACP protein is identified in the invention as		
CC	being megakaryocyte stimulating factor (MSF). The gene encoding		
CC	CACP protein (MSF) is located on chromosome 1q25-31, and mutations in		
CC	this gene are responsible for the heritable disorder campitodactyl-		
CC	arthropathy-coxa vara-pericarditis, in which patients have synovial		
CC	hyperplasia without evidence of inflammation. CACP protein (MSF)		
CC	acts as a synovial lubricant, and can be used to lubricate tissue and		
CC	joints in the treatment of osteoarthritis. The composition may be		
CC	applied to reduce the symptoms of osteoarthritis (e.g., joint pain,		
CC	loss of range of movement or joint damage). The present sequence		
CC	represents human megakaryocyte stimulating factor (MSF, CACP protein).		
CC	Note: This sequence is not given in its entirety in figure 4 of the		
CC	specification, although a GenBank accession number was given. This		
CC	sequence was therefore obtained from GenBank (U70316).		
XX			
XX	Sequence 1404 AA:		
XX			

[illegible]

```

FT      Region          1266..1331
FT      /Label=- Exon_X
FT      Region          1331..1373
FT      /Label=- Exon_XI
FT      Region          1373..1404
FT      /Label=- Exon_XII
XX
XX      MO9213075-A.
PN
PD      06-AUG-1992.
PX
PX      17-JAN-1992;    92MO-US040433.
PE
PR      18-JAN-1991;    91US-0643502.
PR      10-SEP-1991;    91US-0757022.
PA
PA      (GEMV ) GENETICS INST INC.
PI
PI      Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K;
P1
P1      MPI: 1992-284660/34.
DR
DR      N-PADB; AAO272223.
XX
XX
XX      New human mega-karyocyte stimulating factors - for treating
PT      immune deficiencies, cancer, exposure to radiation or drugs,
PT      bacterial and viral infections, etc.
PS
PS      Claim 1, 2 and 3; Fig 1; 87pp; English.
XX
XX
CC      The sequence given is a full length translation from the megakaryocyte
CC      stimulating factor (MSF) precursor. The sequence covered by exons II,
CC      III and IV encodes megakaryocyte stimulating factor (MSP). This
CC      sequence is modified by the addition of an N-terminal sequence encoding
CC      a secretory leader, an initiating methionine preceding exon II and a
CC      terminating codon following exon IV. The cDNA sequence given contains
CC      (meg-CSF). Exon I contains the initiating methionine, and encodes a
CC      classical mammalian protein secretion signal sequence. The sequence
CC      terminate in the region between amino acid residues 134 - 147. The
CC      primary transcript of this gene may be cleaved in different ways to
CC      yield a family of mRNA's each encoding a different MSF protein. Exons
CC      V and VI are thought to be related to the activity of the factor and
CC      are also implicated in the stability, folding and processing of the
CC      molecule. These exons are also thought to play a role in the observed
CC      synergy of MSF with other cytokines. Exons V - XII are believed to be
CC      implicated in the processing or folding of the appropriate structure of
CC      the resulting factor, i.e. one or more of these exons may contain
CC      sequences which direct proteolytic cleavage, adhesion, organisation of
CC      the cellular matrix or extracellular matrix processing. Both naturally
CC      occurring and non-naturally occurring MSF's may be characterised by
CC      various combinations of alternatively spliced exons from this sequence,
CC      with the exons spliced together in differing orders to form different
CC      members of the MSF family.
XX
XX
SQ      Sequence 1404 AA;

Query Match          99.2%; Score 7007.5; DB 13; Length 1404;
Best Local Similarity 93.4%; Pident No. 0;
Matches 1311; Conservative 0; Missmatches 0; Indels 93; Gaps

OY      1 MAMKTLPTYLALLLSVFVIQOVSSODLSSCAAGCEGYSRDATCNDNVNCGHYHMCCEPF 60
DB      |||||
DB      1 mawktlptlylllllefvfIdqvsddlsccagrcgegsrdatcncdyncghymecpdf 60
OY      KRVCFAELSCGRCESEERGEHCDDAQCKKYDKCPDYESFCAE----- 106
DB      |||||
DB      61 krvcfaelscgrcfesfergrecdcagckkydkcpdyesfcaevhnpspskkap 120
OY      ----- 106
DB      121 ppgasqlketkrspkpnpkktkkvileeeelteehsvseengsssssssstliw 180

```

Qy	107	-----VKONKKRRTKKKKPKPPKPPVVDEAGSGJLJNDGDFKVTTJPDST	147
Db	181	kixssksnaaneIqkklkwwkdnkknrckkkbpkpvpvdeagslndgdfkvttptst	247
Qy	148	TOHNKVTSTKITTAKPINDRPLSPNSOTSKEJSLTVNKEJTVTKETTTNNKOTSDG	207
Db	241	tqnmkvsqskltakpnpnpislpnsdtskcselvnkeltvektcttnqtsbdg	307
Qy	208	KEKITSKETOIESTKSAKDLPATSKVLKPKPKAEJTTKGALJTPKEEPTPKPEAS	267
Db	301	kektasaekqslcktsacdIaptskvlakpckpacelttkpaltcpckpctpkcpas	367
Qy	268	TTPKKEPPTTIKAPITPKEPAPITTTKSAPTTPKPEAPITTKEPAPITTKPEAPITTKEP	327
Db	361	tkcpkcpcttkIsapctpkcpapcttkksapctpkcpapcttkcpapcttkcpapcttkcp	427
Qy	328	APPTTKSAPITPKEPAPITPKKPAITPKPEAPITPKPEPTTPPKPEAPITTKPEAPITPK	387
Db	421	apcttksapctpkcpapcttkpkpactpkcpapcttkpkpactpkcpapcttkcpapcttk	487
Qy	388	EPAPITPKKPAITPKPEAPITTKPKPAITTKKESPTTKPEAPITTKSAPITTKPEAPIT	447
Db	481	epctpkpkpactpkcpapcttkcpapcttkcpspttkcpapcttkcsapcttkcpapct	547
Qy	448	TTKSAPITPKESPTTKKEAPITTKKEBAPITTKKPAITTKKEBAPITTKPEAPITTKKP	507
Db	541	tkksapctpkcpspttkcpapcttkcpapcttkpkpactpkcpapcttkcpapcttkcp	607
Qy	508	APAPITPKAPITPKETAAPITTKKLTPTTKPKLAPITPKRAPITTKPEELAPITTKPEEPTPT	567
Db	601	apapkepapcttkpacttkpkltpctpkelapcttkcpapcttkpeelapcttkpeepct	667
Qy	568	PEEPAPITPKAAPITPKKEBAPITTKKEBAPITTKPEAPITTKKEBAPITTKGAPITTKKEB	627
Db	661	peepcttkpkaapcttkpcepcttkpcepapcttkpcepapcttkpcapttkqatcttkcp	727
Qy	628	APPTTKKAPKELAPITTKKEPTSTTSODKPADITTKGAPITTKPEAPITTKPEAPITTKPG	687
Db	721	apcttkpkapkelapcttkcpcttkcsdcpapcttkqatcttkpcepcttkpcepapcttkpg	787
Qy	688	TAPITTKKAPAPITTKPKKPAKELAPITTKGSPSTTSODKPADITTKKEBAPITTKKEBAPITTK	747
Db	781	tapcttkpcepcttkpkpapkkelapcttkkspsttsdcpapcttkpcepapcttkpcepapcttk	847
Qy	748	KPAITTKPTPTPTSEVSTPTTKKEPTTIHKSDEESTPELSAPITKALENSPKKEBGPVT	807
Db	841	kpacttkpctpttksevspttkcpcttklnksdestipelsaepcttkalenspkkebgvpt	907
Qy	808	TKTPATPKPEMTTAKKKTBERDRLRTPETTTAAPKMKETATTTBKTTESKITATTTQV	867
Db	901	tktpacttkpemtctakktleerdrlrttpecttklaapmkctelcttkeskitatctqy	967
Qy	868	TSTTTQDTPPKITTLTKTTLABKVTTTKKTIITTTIIMKKPEETAKPKXKRAANSKATPK	927
Db	961	tscttqdtcpkltlctltcttlapkvcttkctltcttelmkpceetakpkdraensatcpk	1027
Qy	928	POKTPAKPKKPTSTKKKKKTMKPRVRKKTTPTPPKKMTSTMBELNPTSRSLAEAMLQDTPBN	987
Db	1021	poktpakpkpkpctskpkktmpvrkpkcttpckmtstmpelnsptsrslaemlqcttrpn	1087
Qy	988	QTPNSKLEVENPKSEDAAGESTPMMLRLPHVMEFVYPMDDYVLPVPNGIITINPMLS	1047
Db	1081	qtprnsklvevnpsedagsegetcmllrphvfmpevcpndyvlprvnglilnnpmls	1147
Qy	1048	DEHNCKGRVDDGLTLRNGTLYAFNGHFMWLSPPSPSPARRITTEWAGJPSIDVFT	1107
Db	1141	dehncgkpvddgltlrngtlvafngfymwlspspsparrittewgjpsidvftc	1207
Qy	1108	RCNCEGTFFFKSQSQWRTNNDIKDYGKPLTFKFGGJLGOIYVALSTAKYKWPSSVY	1167
Db	1201	rcncegtfffksgywtntndlkdegyjpkplfkfsgjlgclqiyvaalstakymwpssvy	1267

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:25:34 ; Search time 138.84 Seconds
(without alignments)
699.439 Million cell updates/sec

Title: AAB
Perfect score: 7064
Sequence: 1 MAWKPTPIYLLLSVFVIQ.....ARATRRSGTSLSKWYNCP 1311

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101:*

- 1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT:*
- 2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT:*
- 3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT:*
- 4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT:*
- 5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT:*
- 6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT:*
- 7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT:*
- 8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT:*
- 9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT:*
- 10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT:*
- 11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT:*
- 12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT:*
- 13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT:*
- 14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT:*
- 15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT:*
- 16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT:*
- 17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT:*
- 18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT:*
- 19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT:*
- 20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT:*
- 21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:*
- 22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7007.5	99.2	1404	13 AAR26049	MSF precursor. SY
2	7007.5	99.2	1404	22 AAB60568	Human megakaryocyt
3	7007.5	99.2	1404	22 AAB29773	Human megakaryocyt
4	6431.5	91.0	1299	22 AAM24322	Human EST encoded
5	3484	49.3	902	22 AAB29778	Human MSF-derived
6	1934.5	27.4	452	16 AAR80041	Human megakaryocyt
7	1707.5	24.2	472	22 AAB60569	Bovine MSF ortholo
8	1180	16.7	5179	22 AAM24516	C899P predicted am
9	950	13.9	763	21 AAG38942	Arabidopsis thalia
10	950	13.4	1664	19 AAM43106	C. thermocellum Ol
11	776.5	11.0	1325	22 AAM03645	Peptide #2327 enco

12	715	10.1	763	18 AAM31852	Mycobacterium tube
13	627	8.9	4412	21 AAY53666	Sequence g1/101742
14	612	8.7	572	18 AAM31855	Mycobacterium tube
15	547.5	7.8	844	7 AAP60570	Sequence of the Pa
16	542	7.7	807	21 AAY54467	Aminl acid sequenc
17	529	7.5	182	12 AAR10872	Protein encoded by
18	520.5	7.4	788	21 AAY54466	Amino acid sequenc
19	508	7.2	1837	21 AAB11726	Cryptosporidium pa
20	506.5	7.2	744	9 AAR2975	Bioadhesive precut
21	498	7.0	111	13 AAR26050	MSF-K130. Synthet
22	496.5	7.0	2971	21 AAB41231	Human ORFX ORP95
23	489.5	6.9	2972	22 AAB50363	Human SRCAP. Homo
24	489.5	6.9	3118	22 AAB50362	Portion of Cryptos
25	489	6.9	1721	21 AAB11727	P. Yoelli SSP2 ant
26	488.5	6.9	826	13 AAR60042	Peptide #2869 enco
27	488	6.9	617	22 AAM16458	Peptide #2869 enco
28	488	6.9	617	22 AAM04187	Human MUC11 polype
29	488	6.9	957	21 AAY59288	C900U predicted am
30	488	6.9	957	21 AAM48299	Cryptosporidium pa
31	488	6.9	1721	19 AAM48299	Human protein sequ
32	485	6.9	1127	22 AAB95541	Bioadhesive precut
33	476.5	6.7	652	9 AAR2974	Peptide #1317 enco
34	467.5	6.6	511	22 AAM14883	Peptide #1289 enco
35	467.5	6.6	511	22 AAM27312	Peptide #1349 enco
36	467.5	6.6	511	22 AAM02607	Peptide #1289 enco
37	456.5	6.5	750	20 AAY05477	C. albicans Rot1 p
38	450.5	6.4	378	12 AAR14160	PRP 378. Trilicun
39	446.5	6.3	378	12 AAR14162	PRP encoded by clo
40	443	6.3	2870	21 AAY95556	Caenorhabditis ele
41	443	6.3	3178	21 AAY95556	Caenorhabditis ele
42	442.5	6.3	751	16 AAR80839	Japanese sea musse
43	437.5	6.2	2819	12 AAB35408	Human OTCG27 gene
44	429	6.1	1610	21 AAB22870	Human OTCG27 gene
45	429	6.1	1610	21 AAY79648	Human OTCG27 gene

ALIGNMENTS

RESULT	ID	AA	Location/Qualifiers
1	AAR26049	standard; Protein; 1404 AA.	
XX	AC	AAR26049:	
XX	DT	02-FEB-1993 (first entry)	
XX	DE	MSF precursor.	
XX	XX	Megakaryocyte colony stimulating factor; secretion signal; meg-CSF;	
KW	KW	stability; proteolytic cleavage; adhesion; alternative splicing.	
XX	OS	Synthetic.	
XX	FT	Key	
FT	FT	Region	1..26
FT	FT	Region	/label= Exon_I
FT	FT	Region	26..67
FT	FT	Region	/label= Exon_II
FT	FT	Region	67..107
FT	FT	Region	/label= Exon_III
FT	FT	Region	107..157
FT	FT	Region	/label= Exon_IV
FT	FT	Region	157..200
FT	FT	Region	/label= Exon_V
FT	FT	Region	200..1141
FT	FT	Region	/label= Exon_VI
FT	FT	Region	1411..1166
FT	FT	Region	/label= Exon_VII
FT	FT	Region	1166..1212
FT	FT	Region	/label= Exon_VIII
FT	FT	Region	1213..1266
FT	FT	Region	/label= Exon_IX

Mycobacterium tube
Sequence g1/101742
Mycobacterium tube
Sequence of the Pa
Aminl acid sequenc
Protein encoded by
Amino acid sequenc
Cryptosporidium pa
Bioadhesive precut
MSF-K130. Synthet
Human ORFX ORP95
Human SRCAP. Homo
Human SRCAP. Homo
Portion of Cryptos
P. Yoelli SSP2 ant
Peptide #2869 enco
Peptide #2869 enco
Human MUC11 polype
C900U predicted am
Cryptosporidium pa
Human protein sequ
Bioadhesive precut
Peptide #1317 enco
Peptide #1289 enco
Peptide #1289 enco
C. albicans Rot1 p
PRP 378. Trilicun
PRP encoded by clo
Caenorhabditis ele
Caenorhabditis ele
Japanese sea musse
Human OTCG27 gene
Human OTCG27 gene
Human OTCG27 gene
Bifunctional caspa

p. 1-3


```

Db      314 papppapvavplapldisgrvsvwksftlslfcrvcsgvlagalnprrsppl 373
Oy      642 --TTPKAPPTPKKEPAPTPKETA---PTPKGPA---PTTKEPAPTP--KRPAPKE 691
Db      374 ttttclpapldpplpplpplpplntavpplpplpplvltalaplplpplpplspvppap-- 431
Oy      692 LAPPTTKPTSTSDKAPAPTPPKGAPPTPKSPA-----PTTPKKEPAPTPPKGAPPTPL 745
Db      432 --pdpkgkwtclpplapapppk-lvplpplpssppsekmppppppppppppppp 488
Oy      746 KKPAPTPP---KKPAPKEAPPTTKGP--TSTSDKAPAPTPKET---APTTPKEPAPTP 798
Db      489 appapmpsaavpvpsspplpappapaaaprasmpalppappspatrlcpplpsspapasp 548
Oy      799 KKPAPTPPEPTPTTSEVSTPTTKETPTTKHKS---DESPPELSAEPPTPA--LENSP 852
Db      549 --papppapltppklls--anppcpvpapaprrppapppppppalppdpdpvpansp 604
Oy      853 KPPGVPPTTKTPAATKPEMTTTAKDKTTERDLTTPETTTAAKPMKETAATTEKTTESKI 912
Db      605 --papppapppsalpfpvnpa-----pftpaapk-----srpal 637
Oy      913 TATTTQVSTTTQDTTPPKITTLTKTTTLAPKVTTKKTTTTEINKKEETAKPKDRAFN 972
Db      638 paappapppavratp-----ppapppappp 665
Oy      973 SKATTPPKQKP---TKAPKKTSTKPKPKTMPPVKKPKTPTTPPKMTSTM-----PELNP 1023
Db      666 smalpappdpdpplpplapppappppplpmpspappplppaapdpppapplltlmgppspplap 725
Oy      1024 TSRIEALQTTTRNQPNKSLVENVKPSDAGAGCEGT 1063
Db      726 vpgaplapldpplngpfparknsl-----gsssgdt 756

RESULT 13
AAV53666
ID AAV53666 standard; Protein: 4412 AA.
XX
AC AAV53666;
XX
D7 22-FEB-2000 (first entry)
XX
DE Sequence gi/1017427/emb/CAA62189 from an alignment with protein 608.
XX
KW Mechanical stress; gene therapy; protein 608; osteoporosis; bone density;
KN bone development; gi/1017427/emb/CAA62189.
XX
OS Unidentified.
XX
PN WO9960164-A1.
XX
PD 25-NOV-1999.
XX
PE 14-MAY-1999; 99WO-US11066.
XX
PR 15-MAY-1998; 98US-0085673.
XX
PA (OUAR-) QUARK BIOTECH INC.
XX
PI Elinat P, Mor O, Skalliter R, Feinstein E, Faerman A;
XX
DR WPI; 2000-053304/04.
XX
PT Identification of stress induced genes for determining risk and
PT preventing, treating or controlling osteoporosis
XX
PS Claim 32; Fig 6A-R; 308pp; English.
CC
CC The present sequence is obtained from a clustral X alignment with
CC protein 608. Protein 608 was identified using the method of the invention
CC after subjecting rat osteoblasts to mechanical stress. Expression of the

```

```

CC 608 gene was found to be upregulated by about 3-fold in cells subjected
CC to mechanical strain. The specification describes a method for the
CC identification of genes responsive to a specific mechanical stress. The
CC method comprises applying the mechanical stress to an organism (tissue
CC or cells comprising bone cells), isolating the specific cellular
CC fractions and extracting mRNA from them, and differentially analyzing the
CC mRNA in comparison with control samples. The method is used to identify
CC genes whose expression is responsive to a specific stress. The identified
CC genes are employed in determining risk associated with a physiological or
CC disease state. The risk determination methods are used for testing a
CC medicament for gene therapy. These medicaments, or genes identified by
CC the method of the invention, are used for treating, preventing or
CC controlling a physiological or disease state (especially osteoporosis or
CC bone density or other factors causing or contributing to osteoporosis or
CC its symptoms or other conditions involved in mechanical stress or its
CC lack. The methods can also be used for advancing research or studies in
CC bone development.
XX
SQ Sequence 4412 AA;
XX
Query Match 8.9%; Score 645.5; DB 21; Length 4412;
Best Local Similarity 24.0%; Pred. No. 6e-27;
Matches 271; Conservative 128; Mismatches 488; Indels 243; Gaps 52;

Oy      26 ELCKGRCFES-----FERGECDCDOACKYDCCEDYESFCA-----EYHNPSPSS 75
Db      3278 qlschvggsepilrlqwlkage-----lkspdcsfsfsgtvlrlrdvakadsgdyv 3311
Oy      76 KKAPPSGASQTIKSTKRSRPPKPKKTKRVISEE---ITEHVSSENOESSSSSS 132
Db      3332 ckaenavag-sdltckstkltdkdpavatkkaavdgdlffvseqslrvvektatlatiak 3390
Oy      133 SSSSTIKKISSKKSANREL-----QKTLKVKDNKKNKTKKPPKPPVD 179
Db      3391 vvgqdlpvnvkwtkykw--rlnggyrvflhqgdeaklelrdtltk----- 3434
Oy      180 EAGSGLDNGDFKVTTPDSTSTQHNKVTSPKITTAK-----PINRPSLPN 226
Db      3435 -----dsglyrc-----valnehgelsavnlqydekkkgeklegdlramlkkpklkyg 3484
Oy      227 SDTSKETSIVNKEETVEKEETTTNKRGS--TDGKEKTSAKETQSIEKT-----SAK 278
Db      3485 ageeeeldl-mellknvdkreyekyarmygltdfrglqvveehrvckhrylevfeee 3543
Oy      279 DLAPTskVLAKPTPKAETTKGPALE-----TKKEPTPTTP--KEPASTTPKEPTPTTIS 332
Db      3544 everfefekpapp-----kspelskllppkkppkvvprkrepakvpevkkilvvee 3595
Oy      333 APTTPKEP--APTTPKSAPPTPKKEPAPTTTKEPAPTTTKEPAPTTTKSAPTP 391
Db      3596 kvrvpeepvrvpckvpevlppkevvp-----ekvrvppakkpapp-----ppkvp 3641
Oy      392 KEPAPTPPK-----PAPTPPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTP-----KE 440
Db      3642 eapkevvpckvvpvpkvppevplkvpevkaavpekkvpealppkpspppevfeee 3701
Oy      441 PAPAPKKAPAPTTPKKEPAPTTTKEPAPTTTKEPSPTTKEPAPTTTKSAPTTTKEAPPT 500
Db      3702 espapppk-kpevprrvrvpevpekkvpaap--pkkpvevlyk-vpeapkevppk 3757
Oy      501 TKSAPTTTKEPSPTTTKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 560
Db      3758 kvvpvp--pkkkpevppk-----vpevkvavpekkv-----aalppkpspppevfeee 3807
Oy      561 PTAPKEPAPTTTKEPAPTTTTPPKLT-----PTTPEKLAPTT-----PEKRAPTTPEELAPTP 612
Db      3808 evalpeepaeveepeaapppvylvpknvpekkapavavakkelpvkvpevpekvvp 3867
Oy      613 EEPPTP--TTPKEPAPTTPKAAPTKEPAP-----TTPKEP-----APTTPKEPAPTTTPK 661
Db      3868 ekvvpilvvpkk--paappkvpevpevpevpekvavpkkvpevppakvpevppkvp----- 3921

```

```

QY 286 VLAKPTPKAETTTKGPALTTPKEPT-----PTTPKEPASITPKKEPTTTIK 331
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 329 fpaepthegertanenttspaeptehgertanenttspaeptehgertfandktlss 388
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 332 SAPTPPK-EPAPTTTKSAPTTKKEPAPTTTKE-----PAPTTTKEPAPTTTKPA-- 380
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 389 saestehgertplanen--tspaeptenretanenttspaeptenretanektl 446
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 381 ----PT-----TTKSAPTTKKEPAPT-----TPKKPAPTTKKEPAPTTKEPT-----TT 422
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 447 spveptenretanektspaeptengqrfpanektsssaepthegertplanent 506
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 423 PKEPAPPT-TKEPAPTTKKEPAPPAKKPAPTTKKEPAPTTKKEPAPTTKKEPAPTT 481
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 507 ppaepthegertanektspaeptengdrtpfanektspaeptengdrtpfan 563
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 482 APPTTTSA-PTTKKEPAPTTTKSAPTTKKEPAPTTTKE-----PAPTTKKEPAPTT 532
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 564 ektsssaepthegertplanen--tspaeptenretanenttspaeptengqrf 621
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 533 KKRPAPTTKKEPAPTTKKEPAPTTKKRPAPPAKKPAPTTKKEPAPTTKKEPAPTT 592
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 622 nektlfpaeptenretanektsspa--epengqrfpanektsssaepthegert 679
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 593 PTPPEKP-APTPEELAPTTPEEPPTTPEEPAPTTPKAAANTKEPA-PT-----TP 644
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 680 planentlspaeptenretanektspaeptenretanenttspaeptengdrtp 739
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 645 KEPAPTPKKEPAPTT-----TPKETAPTTKGATPTTLEKAPTTPKKAPKELAPTTTKE 699
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 740 lanektspaeptengqrfpanektsssaepthegertplanenttspaeptenre 799
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 700 PTSTSDKPAPTTTPKATPTTTPKEPAPTTKEPAP-----TPKGAAPTTLKAPPA 750
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 800 rtanektspaeptenretanektspaeptenrewtanentlspaeptenremp 859
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 751 TTPKKKAPKELAPTTTGGPTSTSDKPAPTTTPKATPTTTPKEPAP-----TPKKAAPT 805
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 860 lanektspaeptengqrfpanektsssaepthegertplanenttspaepten 915
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 806 PEPPTTSEVPTTTPKEPTTIHKSPESTPELSAEPPTKALENSPKEPGVPTTKTPAA 865
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 916 -ehgerlanekatspkpkehgertfvedtltspaept-----eng-----ertpla 962
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 866 TKPMTTAAKDKTERDLRTTPETTTAAPKMTK-----TATTTKTESKTITATT-TOVT 920
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 963 n-entltspaeptenretanektspaeptehgertspaanektlspakpkeheemt 1020
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 921 STTQODTTPFKI-----TTL-----KTTTLAPKRVTTT----- 947
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1021 psanenttspvxpkehgertfpanektlspaeptehgaktssanektlspakpkehg 1080
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 948 -KKTITTEIMNKPEETAKPKDRATNSKATTPKPKQK-----TK 985
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1081 ertspndktsssaestehgertsaavtltppaepkthakrttlahekmtqytekste 1140
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 986 APKPTSGKKKTKTPRVKRP-----KTTTPRKMTSTMPDELNPTSRTIAEAMLO 1033
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1141 hpektltstte-ktctipekpllyektlctgkntlvpekpentlgtlttelckapvk 1199
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1034 TTTTRNOT 1041
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1200 stenpekt 1207
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 12
AAW31852
ID AAW31852 standard; Protein; 763 AA.

AC AAW31852;
XX
XX
XX 27-APR-1998 (first entry)
XX

```

DE Mycobacterium tuberculosis 74 kDa protein.
XX Tuberculosis; mycobacteria; infection; diagnosis;
XX antimycobacterial; antibiotic; vaccine.
KW Mycobacterium tuberculosis.
OS WO9741252-A2.
PN 06-NOV-1997.
PD 18-APR-1997; 97WO-EP01973.
XX 29-APR-1996; 96DE-4017184.
XX (GBFB ) GBF GES BIOECH FORSCHUNG GMBH.
PA Espilita C, Honisch C, Moreno C, Singh M;
PI WPI: 1997-549750/50.
XX N-PSDB; AAT93610.
DR New DNA and related proteins or RNA derived from M. tuberculosis -
XX used for diagnosis of mycobacterial infections, monitoring
XX vaccination and development of anti-mycobacterial agents
PS Claim 5; Fig 13; 55pp; English.
XX
XX This novel 74 kDa protein is encoded by an open reading frame of
XX a Mycobacterium tuberculosis DNA fragment (see AAT93610) containing
XX a polymorphic GC-rich sequence. Its amino acid sequence shows
XX a high proline content, but there is no homology to any known
XX proline-rich antigens of mycobacteria. Novel M. tuberculosis
XX proteins (see AAW31851-57) are claimed. These can be produced as
XX recombinant proteins, especially in bacterial, yeast, fungal or
XX higher eukaryote host cells, and used for diagnosing tuberculosis
XX and other mycobacterial infections in humans or animals. The
XX claimed proteins can also be used for epidemiological studies, for
XX monitoring vaccination, and for the development of vaccines and
XX anti-mycobacterial drugs.
XX
XX Sequence 763 AA:
SQ

```

Query Match 9.8%; Score 715; DB 18; Length 763;
Best Local Similarity 28.2%; Pred. NO. 1.4e-31;
Matches 248; Conservative 39; Mismatches 361; Indels 232; Gaps 39;

```

QY 290 PTPKAETTTKGPALTTPKEPTTTTPKKEPAPTTTTPKKEPAPTTTTPKKEPAPTTT 349
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3 PVP-----apralaplpapapapapapapapapapapapapapapapapapapapap 46
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 350 TTPKEPAPTTTTPKKEPAPTTTTPKKEPAPTTTTPKKEPAPTTTTPKKEPAPTTTTPKE 409
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 47 pcp--pap-----papkpkskapfpvpapaparelapp--pap-----peapre 90
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 410 PAPTPKEPTPTT-----PKEPAPTTKKEPAPTTTK-EPAPPAK-KKAPTTKKEPAPTTK 463
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 91 srpalpcoppvvpvlpdppeapapvpvppapspfpfpfpapklvppapvp--pvpnspp 148
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 464 EP-APTTPKEPSPPTTTPKEPAPTTTTPKKEPAPTTTTPKKEPAPTTTTPKKEPAPTTT 521
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 149 fbpfpapalnpap-----papplanspplpappfpagc--ppaapwppvpaapkskpa 201
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 522 TTPKEPAPTTTPKKEPAPTTTPKKEPAPTTTTPKKEPAPTTTTPKKEPAPTTTTPKETAP 577
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 202 sprrppap-----pmpatpmeffpplpvpdpdjsketppapppappplpapvpplp 256
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 578 TTPKKLTPPTTPEKLAPE-----TTPKKAAPTTPPEELAPTTTPEEPPTTPEEPAPTTT-- 627
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 257 pvpnkkipap--apvavaavlvapcpplpplpnhppapapavpvyplaplpnshp 313
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 628 ---PKAAPNTPKKEPAP----- 641

```

Oy	345	TKSAPTPKKEBAPTTTKEBAPTTPKEBA----	PTTKEBAPTITTSAPTTPKEBAPTTPK	400
Dd	:	: : : :	: : : :	:
Dd	805	eplrlseepreelprldlpsdeplrsdeplrsdeplrsdeplrsdeplrsdeplrsdeplrlclp		864
Oy	401	KPAPTTPKEBAPTTPKPRTPTTPKEBAPTTPKEBAPT--TPKEBAPTAPKKAPATTTPKEBAP		459
Dd	:	: : : :	: : : :	:
Dd	865	sdeplrsdeplrs--deplrs--depltp-sdeplrsdeplrlclpsdeplrsdeplrtcp		919
Oy	460	TTPKEBA--PTTTPKEBAPT--TPKEBAPTITTKS--APTTPKEBAPT---	TTKSAPTTPKEBS	512
Dd	:	: : : : : :	: : : : : :	:
Dd	920	sdeplrsdeplrsdeplrsdeplrsdeplrlclpsdeplrsdeplrsdeplrsdeplrsdeplrt		979
Oy	513	-----PTTTPKEBAPT--TPKEBAPTTPKKRPAATTTPKEBAPTTPKEBAPTITTKAKAPAPKEP		567
Dd	:	: : : : : :	: : : : : :	:
Dd	980	psdeplrsdeplrsdeplrsdeplrlclpsdeplrsdeplrsdeplrsdeplrsdeplrsdeplrsdep		1039
Oy	568	APT---TPKENAPTTPPKLPTTPPEKLAPTTPKEBAPT---	PPEELAPTTPPEEPT--TP	620
Dd	:	: : : : : :	: : : : : :	:
Dd	1040	tpsdeplrspe-----tlseplrlclpsdeplrsdeplrsdeplrsdeplrt--eplrsdeplrssetp		1094
Oy	621	EEPAPTTPKAABNTTPKEBAPTTPKEBA--PTTTPKEBAPT--TPKENAPTTPKGATPTTLK		677
Dd	:	: : : : : :	: : : : : :	:
Dd	1095	eeprlclpsdeplrsdeplrsdeplrsdeplrsdeplrsdeplrsdeplrsdeplrlclpsdeplrsd		1154
Oy	678	EPAPTTPKKKAPKELAPTTTKERT--STSDKAPATTTPKGAATPTTPKEBAPTTPKKBAPT		736
Dd	:	: : : : : :	: : : : : :	:
Dd	1155	eplrs--deplrsd-eplrsdeplrsdeplrsdeplrlclpsdeplrsdeplrs--deplrs-		1208
Oy	737	PKGAPTPTTLKEBAPT--PKKPAPKELAPTTTKGSTSDKAPAPT---	TPKETAPTTPKE	792
Dd	:	: : : : : :	: : : : : :	:
Dd	1209	----deplrsdeplrsdeplrsdeplrt-----ptlrsdeplrsdeplrsd-eplrsde		1253
Oy	793	PAPTTPKKRAPT--TPETTPPTTSEVSTPTTTPKEBAPTTHKSPDESJPELSAEPTRALENS		851
Dd	:	: : : : : :	: : : : : :	:
Dd	1254	ptps--deplrsdeplrsdeplrlclpsdeplrsdeplrt-----psdeplrs--sdeplrsdeplrt		1305
Oy	852	PKEGCVTTTKTPAAT--KPEMTTAADKTTERDLRTTEPETTTAAKPMKETAJTTEKTE		909
Dd	:	: : : : : :	: : : : : :	:
Dd	1306	sdeplrsdeplrsdeplrsdeplrlclpsdeplrsd-----eplrsdeplrsdeplrsdeplrs		1361
Oy	910	SKIRATTTQVTSTTOOTTPFKITTLTKTTTLAKAVTTTKKITTETEIMNKPBEAKPKDR		969
Dd	:	: : : : : :	: : : : : :	:
Dd	1362	deplrsdeplrsdeplrsdeplrcp-----tlscplrsclt-----	psg	1396
Oy	970	ATNSKAT-----TPKPOKPTPKAP--KKPTSKKKKTTPRAVRKPKXTPTPRKM-----		1014
Dd	:	: : : : : :	: : : : : :	:
Dd	1397	sggsggsgggsgggsggvrvpsrclpsrclpskxtstap--tealeerpdsavpaigshra		1453
Oy	1015	-----STMPELNPTSRIAEMLQOTTTRPNOTPNKSLVEYNPKSEDAGAEGETPHM-		1066
Dd	:	: : : : : :	: : : : : :	:
Dd	1454	yrlgydpdgstfrpernltrleaaavlf-----akll-----gadesygagssapsyd		1498
Oy	1067	LLRPH-----VEMPE-----VTDDMOVLPRVPNOGI-----		1092
Dd	:	: : : : : :	: : : : : :	:
Dd	1499	ladchwaawaikfatagsglkyypdgtkfkpdgnittraefatvlnflitkvxqgeImksla		1556
Oy	1093	--IINPLSDETNICNG---RPVGDLTTL-----RGTGLVAFRGHFTMMLSFSSPPS		1139
Dd	:	: : : : : :	: : : : : :	:
Dd	1559	tldslnpxfdd---cvghwagefielhlslylgyspygd-----fkpgn		1600
Oy	1140	PARATEYWGIPSPIDIVTFTRCMCEGKTFFEKD--SOYKRFPNDIKD		1184
Dd	:	: : : : : :	: : : : : :	:
Dd	1601	yikrseesv---alInratelrglnabaplfpbvnesyaaf-gdlmd		1642

RESULT	11
AA003645	
XX	AA003645 standard; Protein; 1325 AA
AC	AA003645;
XX	
DT	09-OCT-2001 (first entry)

DE Peptide #2327 encoded by probe for measuring breast gene expression.
XX
XX Probe; human; breast disease; breast cancer; development disorder;
KV inflammatory disease; proliferative breast disease; non-carcinoma tumour.
KW
XX
OS Homo sapiens.
XX
XX WO200157270-A2.
PN
XX
PD 09-AUG-2001.
XX
XX
PF 29-JAN-2001; 2001WO-US00561.
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-476286/51.
DR
XX
XX Novel single exon nucleic acid probe used to measuring gene expression
PT In a human breast -
PI
XX
XX
PS Claim 27; SEQ ID NO 12385; 322pp; English.
XX
XX
CC The present invention relates to novel single exon nucleic acid probes
CC (see ARI00010-ARI10067). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for measuring human gene expression in
CC a human breast sample, where the probe hybridises at high stringency to a
CC nucleic acid expressed in the human breast. The probes are useful for
CC predicting, diagnosing, grading, staging, monitoring and prognosing
CC diseases of the human breast, particularly those diseases with polygenic
CC aetiology. The diseases include; breast cancer, disorders of development,
CC inflammatory diseases of the breast, fibrocystic changes, proliferative
CC breast disease and non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 1325 AA;

Query Match	Similarity	Score	ID	Length
Best Local	26.2%	Pred. No. 3.1e-36		
Matches	317	Conservative	122	Mismatches 517; Indels 252; Gaps 47
QY	35	ESFEGREDCDCAOACKKYKCCP-----DSEFCAEVHNPTSPSSKRAAPPSCASOT	87	
DB	51	ehnegkhcnhtlrhskprdk-ptgnsktidhks---stcdhnapptlseansngkdkpm	105	
QY	88	IKSTIKRSKPPNNKKTKKIVIESEELTEHSHSVSENDSSSSSSSSSTIKIKSSKNS	147	
DB	106	ltn--qrsvpdadstltnk-----esagkhhilpaprsklncrksltvgksltv-----trksd	155	
QY	148	AAANRELOKKLKYKD-----NKKNRNTKKRPPKPPVVDVDESGLDNDFV	192	
DB	156	ktgprlekamsldkstsahktctsfhngsqstqkqsktsfpekltaa-----sktylt	211	
QY	193	T--TPDSTQHNKVSF--SPKITTAKPIPRSLPN-----SDRSKETSLVNNKFTV	243	
DB	212	tycpeseketserltvasadkllktctkridgelisanelgtslaeplehggrtlanenmp	271	
QY	244	ETKETTTTNKQSTDOKEKTT--SAKETQSIEKTSAKDLAP-----TSK	285	
DB	272	saepelennerla---neutllispeepetrenetlennatfaapbplennantentll	328	

PR	30-APR-1999,	9905-0132407
PR	04-MAY-1999,	9905-0132484
PR	05-MAY-1999,	9905-0132485
PR	06-MAY-1999,	9905-0132486
PR	07-MAY-1999,	9905-0132487
PR	11-MAY-1999,	9905-0132863
PR	14-MAY-1999,	9905-0134256
PR	14-MAY-1999,	9905-0134218
PR	14-MAY-1999,	9905-0134221
PR	18-MAY-1999,	9905-0134370
PR	19-MAY-1999,	9905-0134768
PR	20-MAY-1999,	9905-0134941
PR	21-MAY-1999,	9905-0135124
PR	24-MAY-1999,	9905-0135353
PR	25-MAY-1999,	9905-0135621
PR	27-MAY-1999,	9905-0136302
PR	28-MAY-1999,	9905-0136389
PR	01-JUN-1999,	9905-0136787
PR	03-JUN-1999,	9905-0137222
PR	04-JUN-1999,	9905-0137528
PR	07-JUN-1999,	9905-0137502
PR	08-JUN-1999,	9905-0137724
PR	10-JUN-1999,	9905-0138094
PR	10-JUN-1999,	9905-0138540
PR	14-JUN-1999,	9905-0138847
PR	16-JUN-1999,	9905-0139119
PR	16-JUN-1999,	9905-0139452
PR	17-JUN-1999,	9905-0139454
PR	18-JUN-1999,	9905-0139455
PR	18-JUN-1999,	9905-0139456
PR	18-JUN-1999,	9905-0139457
PR	18-JUN-1999,	9905-0139458
PR	18-JUN-1999,	9905-0139459
PR	18-JUN-1999,	9905-0139460
PR	18-JUN-1999,	9905-0139461
PR	18-JUN-1999,	9905-0139463
PR	18-JUN-1999,	9905-0139465
PR	18-JUN-1999,	9905-0139760
PR	21-JUN-1999,	9905-0139763
PR	22-JUN-1999,	9905-0139817
PR	23-JUN-1999,	9905-0139899
PR	23-JUN-1999,	9905-0140353
PR	24-JUN-1999,	9905-0140354
PR	28-JUN-1999,	9905-0140655
PR	29-JUN-1999,	9905-0140823
PR	30-JUN-1999,	9905-0140991
PR	01-JUL-1999,	9905-0141287
PR	01-JUL-1999,	9905-0141842
PR	02-JUL-1999,	9905-0142154
PR	02-JUL-1999,	9905-0142250
PR	06-JUL-1999,	9905-0142395
PR	08-JUL-1999,	9905-0142803
PR	09-JUL-1999,	9905-0142820
PR	12-JUL-1999,	9905-0142977
PR	13-JUL-1999,	9905-0143342
PR	14-JUL-1999,	9905-0143624
PR	15-JUL-1999,	9905-0144005
PR	16-JUL-1999,	9905-0144085
PR	16-JUL-1999,	9905-0144086
PR	19-JUL-1999,	9905-0144325
PR	19-JUL-1999,	9905-0144331
PR	19-JUL-1999,	9905-0144332
PR	19-JUL-1999,	9905-0144333
PR	19-JUL-1999,	9905-0144334
PR	20-JUL-1999,	9905-0144335
PR	20-JUL-1999,	9905-0144352
PR	20-JUL-1999,	9905-0144632
PR	21-JUL-1999,	9905-0144684
PR	21-JUL-1999,	9905-0144814
PR	21-JUL-1999,	9905-0145086

PR	21-JUL-1999;	9905-0145088
PR	22-JUL-1999;	9905-0145085
PR	22-JUL-1999;	9905-0145087
PR	22-JUL-1999;	9905-0145089
PR	23-JUL-1999;	9905-0145143
PR	23-JUL-1999;	9905-0145145
PR	23-JUL-1999;	9905-0145158
PR	26-JUL-1999;	9905-0145224
PR	27-JUL-1999;	9905-0145276
PR	27-JUL-1999;	9905-0145913
PR	27-JUL-1999;	9905-0145919
PR	28-JUL-1999;	9905-0145951
PR	02-AUG-1999;	9905-0146386
PR	02-AUG-1999;	9905-0146388
PR	02-AUG-1999;	9905-0146389
PR	04-AUG-1999;	9905-0147038
PR	04-AUG-1999;	9905-0147204
PR	04-AUG-1999;	9905-0147302
PR	05-AUG-1999;	9905-0147192
PR	06-AUG-1999;	9905-0147260
PR	06-AUG-1999;	9905-0147303
PR	09-AUG-1999;	9905-0147416
PR	09-AUG-1999;	9905-0147493
PR	10-AUG-1999;	9905-0147935
PR	11-AUG-1999;	9905-0148171
PR	12-AUG-1999;	9905-0148319
PR	13-AUG-1999;	9905-0148341
PR	13-AUG-1999;	9905-0148565
PR	16-AUG-1999;	9905-0148684
PR	17-AUG-1999;	9905-0149368
PR	18-AUG-1999;	9905-0149175
PR	20-AUG-1999;	9905-0149722
PR	20-AUG-1999;	9905-0149723
PR	23-AUG-1999;	9905-0149929
PR	23-AUG-1999;	9905-0149902
PR	25-AUG-1999;	9905-0149930
PR	26-AUG-1999;	9905-0150566
PR	26-AUG-1999;	9905-0150884
PR	27-AUG-1999;	9905-0151065
PR	27-AUG-1999;	9905-0151066
PR	30-AUG-1999;	9905-0151080
PR	31-AUG-1999;	9905-0151303
PR	01-SEP-1999;	9905-0151438
PR	07-SEP-1999;	9905-0151930
PR	10-SEP-1999;	9905-0152363
PR	13-SEP-1999;	9905-0153070
PR	15-SEP-1999;	9905-0153758
PR	16-SEP-1999;	9905-0154018
PR	20-SEP-1999;	9905-0154039
PR	22-SEP-1999;	9905-0154779
PR	23-SEP-1999;	9905-0155149
PR	24-SEP-1999;	9905-0155466
PR	28-SEP-1999;	9905-0155659
PR	28-SEP-1999;	9905-0156438
PR	29-SEP-1999;	9905-0156596
PR	04-OCT-1999;	9905-0157117
PR	06-OCT-1999;	9905-0157665
PR	07-OCT-1999;	9905-0158029
PR	08-OCT-1999;	9905-0158232
PR	12-OCT-1999;	9905-0158366
PR	13-OCT-1999;	9905-0159293
PR	13-OCT-1999;	9905-0159294
PR	14-OCT-1999;	9905-0159525
PR	14-OCT-1999;	9905-0159530
PR	14-OCT-1999;	9905-0159531
PR	14-OCT-1999;	9905-0159637
PR	18-OCT-1999;	9905-0159638
PR	21-OCT-1999;	9905-0159584
PR	21-OCT-1999;	9905-0160741

[illegible]

PI Gu X, Han Z, Shen Q:
XX
XX WPI: 1995-330576/41.
DR N-PSDB: AAT04546.
XX
XX
PT New hematopoietic cell growth factor - used for treating
thrombocytopenia and hematocytopenia
PS
PS Example; Page 23; 36pp; Chinese.
XX
XX This sequence represents the human megakaryocytopoietin (MPO) protein.
CC This sequence was purified using a carrier which can couple wheat germ
CC agglutinin and heparin to separate MPO. Fragments of this sequence (see
CC AAR80039 and AAR80040) were used to produce the amplification primers
CC shown in AAT04544 and AAT04545. The fragments amplified by these primers
CC can then be used as probes to screen human CDNA libraries for MPO CDNA.
CC The MPO CDNA can then be inserted into a plasmid which is used to
CC transform cells to produce MPO. The MPO sequence is capable of promoting
CC colony formation of megakaryocytes, enlarging the size of megakaryocytes
CC and stimulating the proliferation of multipotential stem cells. The
CC factor may be used for treating thrombocytopenia and hematocytopenia.
CC The purification method can be used to isolate MPO from human urine or
CC serums of patients with aplastic anaemia, and from animal blood or urine
CC by radiation exposing the animals to induce aplastic anaemia.
XX
XX
SQ Sequence 452 AA;

Query Match	26.3%;	Score 1916.5;	DB 16;	Length 452;
Best Local Similarity	76.1%;	Pred. No. 1.3e-97;		
Matches 378;	Conservative	3;	Mismatches 21;	Indels 95;
				Gaps 33;

QY	1	MAKMTPLVILLLLSVVIVQVSSQ-----	25
Db	1	maMTPLdlyllllsvtlqvsqdlsscagrtogegysrdalcncdyncqhmecpaf	60
QY	26	-----ETSCGRCFESFEGREDCDAQCCKYDKCCPDIESPCAEVHNPTSPSSKKAP	79
Db	61	krcvtaelsckgicrfesfeergredcdagckkydkccpdyfsefcaeivhnptspssk	120
QY	80	PPSGASQTIKSTTKRSPKPPNKKTKKKVIESELTTEHVSSENOESSSSSSSSSSSTIM	139
Db	121	ppsgasqtlketkrspkppnkkkkkkvieselte-----	156
QY	140	KIKSSKNSAANRELQKLIKAKDNKKNKTTKKPTPKPVYVDAGSGLDNGDFKYTPPTDS	199
Db	157	-----yadnknkntkkkprtkprrvdaagsglangdfkvtcpdst	197
QY	200	TOHNKVSIPKTIITAKPINPPSPSPNSDSIKETSLVNKKEVTEPKETTTTKKQSTDS	259
Db	198	tqhmvsvspslitaakrpnppslppnsdskesilvknkeltvetkettlttkqgstdg	257
QY	260	KEKTTSAKETQSIKETSADKLAPTSKVLAKPPTKAEITTTKGPALTTPKEPPTPKEPAS	319
Db	258	kektsaetqsietsakdlaptskvlakrptkpaelttkgpaltpckepptpkcpas	317
QY	320	TTPKPEPPTTIKSAPATTPKEPAPPTTYSAPPTPKPEAPPTTKKEPAPPTTKKEP	379
Db	318	ttpkcppttlksapattpkcpattkcsapltkcpaplttkcpapltkcpaplttkcp	377
QY	380	APTITKSAPPTPKPEAPPTPKKAPPTTKKEPAPPTPKPEPPTTKPKPEAPTTK	439
Db	378	aplttkshpprlprscxxxxctqp-----tkpckpnp prslhprtkcpapltk	426
QY	440	BPAPTAPKKPAPPTPKPE	456
Db	427	epapcarkkppap pple	443
RESULT	7		
ID	AAB60569		
	AAB60569 standard; Protein: 472 AA.		


```
|||||
Db 721 aptlpkppakpkelaplttkeptsstdskpaptlpkgtaptlpkeaplttkpkeaplttkp 780
Oy 740 TAFTLKKEPAPTTPKKPAKELAPTTTKGPTSTSDKAPATTPPKELAPTTKPEAPATTPK 799
Db 781 taplttkpkeaplttkpkeaplttkgptstsdskpaptlpkgtaptlpkeaplttkpkeaplttkp 840
Oy 800 KPAPTPEPTPEPTSEVPTTTEKPEPTTIHKSPESTPELSAEPPTKALENSPKESGVP 859
Db 841 kpaaptpeppteptsevpitsekpttketltkhspestdelsaepitkalsenpkpkegvp 900
Oy 860 TKTPAATKPEMTTAKDKTTERDLRTPEPTTAAPKMTKETAATTTKTESKTTATTVOY 919
Db 901 tktpaatkpemttakdkkterdlrtpepttaapkmktetalttktleskltatttgy 960
Oy 920 TSTTTQDTTPFKTTTLTKTTLAKVTTTKTTTTEIMNKPEETAKPKDRTSKATTPK 979
Db 961 tstitqdtlpfkltltkttllakvlttkltteteimnkpeetapkdatatskatttkp 1020
Oy 980 POKPTAKPKPTSTKKPKTMPRVKRPKPTPEPTPKKMTSTMBELNPTSRIAEAMLQTTTRPN 1039
Db 1021 poptkapkptsttkkpktmprvrkpttptrtkmtsmbelnptsrilaeamlqtttrpn 1080
Oy 1040 QTENSKLVEVNPKSEDAAGAEGETPHMLLRPHVMEPEVPTDMOYLPRVPOGIINPMLS 1099
Db 1081 qtpnslkvevnpksedaagaegelpmllrphvmepevptdmoylprvpogiiinpmis 1140
Oy 1100 DEFNICKGRVVDGLTTLRNGLTVAFRGHYFWMLSPESSPSPARITTEWMCISPIDTFT 1159
Db 1141 dechnicgkvvdglttlrlngltvafrghyfwmlspfspparrittevwgispidftvt 1200
Oy 1160 RCNCEGTEFFPKDSQVYRFTNDIKDAGYPRKRGFGGLTGOYVAAIATKAYKNWPSVY 1219
Db 1201 rcncegtrffkdsqvyrfndikdagypkrlkgilgqilvaalstakynwpsvy 1260
Oy 1220 PFKRGSIIOOYIYKOEVOVOCOPRRPALANYPVGEVMTQVRRRPERAIGPSQHTTIRIQY 1279
Db 1261 ffrkggsiiooyiykoevpocoprrpalanypvgemqvvrrrferaisgsqhtlriqy 1320
Oy 1280 SPARLAVODKGVLEHNEKVSILIMRGLPNVVTSAISLPIKRPDGYDYAASKOQYVINDY 1339
Db 1321 sparlaydkgvlnhevksllwrglpnvvtalslpiirkpdydyatskdyynldv 1380
Oy 1340 PSRTARAITTSSGOTLSKWNCP 1363
Db 1381 psrtaraittrsgqlskwnp 1404

RESULT 4
AAM24322
ID AAM24322 standard; Protein: 1299 AA.
AC XX
AAM24322:
DE 12-OCT-2001 (first entry)
DX XX
Human EST encoded protein SEQ ID NO: 1847.
XX XX
Human, sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
KW biodiversity; gene therapy; nutrition.
OS Homo sapiens.
XX XX
PN WO200154477-A2.
PD 02-AUG-2001.
XX XX
25-JAN-2001; 2001WO-US02687.
XX XX
25-JAN-2000; 2000US-0491404.
PR 17-JUL-2000; 2000US-0617746.
PR
```

```
PR 03-AUG-2000; 2000US-0631451.
PR 15-SEP-2000; 2000US-0663870.
XX XX
PA (HYSE-) HXSEQ INC.
PI Tang YF, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Wehman T;
DR WPT: 2001-476164/51.
DR N-PSDB: AAH98981.
XX XX
PT Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use.
PS Claim 20; Page 1198-1201; 1275pp; English.
XX XX
CC The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a
CC protein of the invention.
XX XX
SQ Sequence 1299 AA;
```

```
Query Match 91.7%; Score 6669.5; DB 22; Length 1299;
Best Local Similarity 96.8%; Pred. No. 0;
Matches 1254; Conservative 0; Mismatches 0; Indels 41; Gaps 1;
```

```
Oy 1 MAMKTLPIYLLLSYFVLIQOVSSQ----- 25
Db 1 mawktlpiylylllsyfvliqovssqqlsacagrcgyardatcncdyqhyamecpdf 60
Oy 26 -----ELSCGKCFESFERGRCDCDAQCKYDKCCPDYESCAVHNPTSPSSKKAP 79
Db 61 kvctaelscgkcfesfergrecdcdaqckkydkccpdyesfcaevhnptspsskkap 120
Oy 80 PRGASQYTIKSTKRSPKPNKKTKKVTSEETHEHSVENQESSSSSSSSSSSTIWM 139
Db 121 prgsaqytlstkrspkpnkktkkvtseethehsvenqessssssssstlwm 180
Oy 140 KIKSSKNANRELOKKLAKVKNKNNKTKKTPKPPVYDEAGSGLDNGDFKYTPDTST 199
Db 181 kiksksnanrelgkklkvknnkntkktpkppvydeagsgldngdfkytptdst 240
Oy 200 TQHNKVTSPKITTAKPILNRPSPSPNDSOTSKESLTVNKETVETKETTNTKOTSDG 259
Db 241 tqhnkvstspkittakpilnrpslpsnpsdskesltvnmkettvetcttnkgqstldg 300
Oy 260 KEKTSKAKETOSTIKTSKDLAPTSKVLAKPTPKAETTTGPAITTPKPEPTTPKPEAS 319
Db 301 kektsaketosstiktskdlaptskvlakptpkasetttgpaalttkpkepttkpepas 360
Oy 320 TTPEKEPTTTIKSAPTTPKKEPAPTTPKSAPTTKKEPAPTTPKKEPAPTTPKKEP 379
Db 361 ttpekeptttiksapttpkkepapttkpsaptttkkepapttkkepapttkkep 420
Oy 380 APTTTKSAPTTPKRPAPATTPKRPAPATTPKRPAPATTPKRPAPATTPKRPAPATTPK 439
Db 421 aptttksapttpkrapattpkrpapattpkrpapattpkrpapattpkrpapattpk 480
Oy 440 EPAPTAPKRPAPATTPKPEAPATTPKPEAPATTPKPEAPATTPKPEAPATTPKPEAP 499
Db 481 epaptapkrpapattpkpeapattpkpeapattpkpeapattpkpeapattpkpeap 540
Oy 500 TTKSAPTTPKESPPTTKKEPAPTTPKKEPAPTTPKRPAPATTPKKEPAPTTPKKEP 559
Db 541 tksapttpkesppttkkepapttkkepapttkkrpapattpkkepapttkkepapttkp 600
Oy 560 APTAPKRPAPATTPKETAATTPKKLPTTPPEKLAPTTPEKRPAPATTPPEELAPTTPEEPTT 619
```

QY 980 POKPTAKPKPTSTKPKTKMPVRKPTPTPRKMTSTMPELNPSRIAEMLQTTTRPN 1039
 DB 1021 POKPTAKPKPTSTKPKTKMPVRKPTPTPRKMTSTMPELNPSRIAEMLQTTTRPN 1080
 QY 1040 QFNNSLVENPKSEDAEGEPTMLLRPRVMEPTVPMDDYLPRVPMGIIITNPMIS 1099
 DB 1081 QFNNSLVENPKSEDAEGEPTMLLRPRVMEPTVPMDDYLPRVPMGIIITNPMIS 1140
 QY 1100 DETNINCKRPGVDGLTLRNGTLVAFRGHYFWMLSPSPSPARRITTEWIGISPIDVFT 1159
 DB 1141 DETNINCKRPGVDGLTLRNGTLVAFRGHYFWMLSPSPSPARRITTEWIGISPIDVFT 1200
 QY 1160 RGNCEGKPEEFKDSQYWRNTNINRAGTPKPFKFGGLTGOIYALSTAKYKNMESVY 1219
 DB 1201 RGNCEGKPEEFKDSQYWRNTNINRAGTPKPFKFGGLTGOIYALSTAKYKNMESVY 1260
 QY 1220 FPKRGSIDQYLYKQEPVOKCPGRPPALNYPYGMTQVRRRRRFRRAIGSGQTHIRIOX 1279
 DB 1261 FPKRGSIDQYLYKQEPVOKCPGRPPALNYPYGMTQVRRRRRFRRAIGSGQTHIRIOX 1320
 QY 1280 SPARLAYODKGYLHNEVKVSIIMRGLPNVVTSAISLPNIRKPDGYDYAFSKDQYTNIDV 1339
 DB 1321 SPARLAYODKGYLHNEVKVSIIMRGLPNVVTSAISLPNIRKPDGYDYAFSKDQYTNIDV 1380
 QY 1340 PSRTARATITRSGOTLSKVMYVNC 1363
 DB 1381 PSRTARATITRSGOTLSKVMYVNC 1404

RESULT 3

AAB29773
 ID AAB29773 standard; Protein: 1404 AA.

AC AAB29773;

DT 28-FEB-2001 (first entry)

XX Human megakaryocyte stimulating factor (MSF), SEQ ID NO:1.

XX Human MSF, megakaryocyte stimulating factor; tribonectin;

KW alternative splicing; joint boundary lubricant; O-linked oligosaccharide;

KM osteoarthritis; tribonectin; tissue adhesion inhibition;

KW friction coefficient reduction; gene therapy; antiarthritic;

XX osteopathic.

XX Homo sapiens.

OS Homo sapiens.

PN MO200064930-A2.

XX 02-NOV-2000.

PD 24-APR-2000; 2000MO-US10953.

XX 23-APR-1999; 9905-0298970.

XX (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.

XX Jay GD;

PI MPI: 2001-024673/03.

DR N-PSDB; AAC81498.

XX Novel tribonectin polypeptide useful as lubricant for treating
 PT osteoarthritis, comprises O-linked lubricating moiety
 PS Claim 3; Page 7; 47pp; English.
 CC The invention relates to a human tribonectin which is a product of
 CC alternative splicing of the human MSF (megakaryocyte stimulating factor)
 CC gene. The tribonectin has at least one O-linked oligosaccharide
 CC lubricating moiety and has a polypeptide sequence comprising 1-76
 CC repeats of a motif having at least 50% identity to the sequence KERPPTT
 CC (AAB29774). The invention also relates to a nucleic acid encoding a

CC human MSF-derived tribonectin; a biocompatible composition comprising a
 CC human tribonectin for inhibiting tissue adhesion formation; and a method
 CC of diagnosing osteoarthritis or a predisposition to osteoarthritis by
 CC measuring the amount of MSF or its fragment in a biological sample of a
 CC mammal, wherein an increased amount of MSF compared to a control
 CC indicates the presence of or predisposition to developing
 CC osteoarthritis. The tribonectin and DNA encoding it are useful in the
 CC treatment of osteoarthritis, where they may be used for lubricating
 CC mammalian joints, such as articulating joints of humans, dogs or horses.
 CC The tribonectin, when formulated as a membrane, foam, gel or fibre, is
 CC useful for inhibiting adhesion between two surfaces such as the injured
 CC tissues of a mammal, where the injury is caused by a surgical insertion
 CC or trauma, or an artificial device e.g., an orthopaedic implant. In
 CC particular, one of the surfaces is pericardial tissue. DNA encoding a
 CC tribonectin may be used in gene therapy. The present sequence represents
 CC human MSF.

Sequence 1404 AA:

Query Match 99.68; Score 7245.5; DB 22; Length 1404;
 Best Local Similarity 97.18; Pred. No. 0;
 Matches 1363; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY 1 MAMKTLPIYLLILSVFVIOQVSQ----- 25

DB 1 MAMKTLPIYLLILSVFVIOQVSQ----- 25

QY 26 -----ELCKGRCFESFERGECDDAQCKKYDKCCPDYESFCALVHNPTSPSSKKAP 79

DB 61 KRYCTAELSCJRCFESFERGECDDAQCKKYDKCCPDYESFCALVHNPTSPSSKKAP 120

QY 80 PPSGASOTIKSTKRSPKPNKKKKKVVISESEITEHSEHSENOBSSSSSSSSSTTW 139

DB 121 PPSGASOTIKSTKRSPKPNKKKKKVVISESEITEHSEHSEHSEHSEHSEHSEHSEHSEH 180

QY 140 KIKSSKNSAANRELQKLLKYNKKNNRTKKKPPKPPVDEAGSGDNGDFKTPPTDTS 199

DB 181 KIKSSKNSAANRELQKLLKYNKKNNRTKKKPPKPPVDEAGSGDNGDFKTPPTDTS 240

QY 200 TQHNKYSTIKITAPINRSLPNSDTSKITSLTVNKETTVEKETTINKOTSTG 259

DB 241 TQHNKYSTIKITAPINRSLPNSDTSKITSLTVNKETTVEKETTINKOTSTG 300

QY 260 KEKTSAKETQSILEKTSANDLAPTSTKVLAKPTPKAETTTGPAITTPKPPPTTPPEPAS 319

DB 301 KEKTSAKETQSILEKTSANDLAPTSTKVLAKPTPKAETTTGPAITTPKPPPTTPPEPAS 360

QY 320 TTPKEPTPTTIKSAPTTPKEBAPTTPKSAPTTPKEBAPTTPKSAPTTPKEBAPTTPK 379

DB 361 TTPKEPTPTTIKSAPTTPKEBAPTTPKSAPTTPKEBAPTTPKSAPTTPKEBAPTTPK 420

QY 380 APTTASAPTTPKEBAPTTPKSAPTTPKEBAPTTPKSAPTTPKEBAPTTPKSAPTTPK 439

DB 421 APTTASAPTTPKEBAPTTPKSAPTTPKEBAPTTPKSAPTTPKEBAPTTPKSAPTTPK 480

QY 440 EPAPTPAKKRPAPTPKEBAPTTPKSAPTTPKEBAPTTPKSAPTTPKEBAPTTPK 499

DB 481 EPAPTPAKKRPAPTPKEBAPTTPKSAPTTPKEBAPTTPKSAPTTPKEBAPTTPK 540

QY 500 TTKSAPTTPKEBAPTTPKSAPTTPKEBAPTTPKSAPTTPKEBAPTTPKSAPTTPK 559

DB 541 TTKSAPTTPKEBAPTTPKSAPTTPKEBAPTTPKSAPTTPKEBAPTTPKSAPTTPK 600

QY 560 APTAPKEBAPTTPKSAPTTPKEBAPTTPKSAPTTPKEBAPTTPKSAPTTPKEBAPTTPK 619

DB 601 APTAPKEBAPTTPKSAPTTPKEBAPTTPKSAPTTPKEBAPTTPKSAPTTPKEBAPTTPK 660

QY 620 PEEBAPTTPKAAAPTPKEBAPTTPKSAPTTPKEBAPTTPKSAPTTPKEBAPTTPK 679

DB 661 PEEBAPTTPKAAAPTPKEBAPTTPKSAPTTPKEBAPTTPKSAPTTPKEBAPTTPK 720

QY 680 APTTPKAPKAPKELAPTTKEPTSTSDKAPAPTTGAPTTPKBAPTTPKEBAPTTPK 739

OY	1220	FFKRGSGIQQIYITTKOEPYQKCPGRPALNTPYGGMTQYRRRRFRRAIGPSOTHTIRIOY	1279
OY	1261	FFKRGSGIQQIYITTKOEPYQKCPGRPALNTPYGGMTQYRRRRFRRAIGPSOTHTIRIOY	1320
OY	1280	SPARLAYQDKGVLHNEVKVSIILMRGLPNVVTSAISLPNTRKPDGYVAFSKDOYVNDIV	1339
Db	1321	sparayqdkgvlhnevksilwrglpmvvttsaislpntrkpdgydyafskdyynldv	1380
OY	1340	PSRTARAITTRSGQTLRSKWNCP	1363
Db	1381	psrtaraittrsgqtlrskwncp	1404
RESULT	2		
AAB60568	ID	AAB60568 standard; Protein; 1404 AA.	
XX	AC	AAB60568;	
XX	DT	27-APR-2001 (first entry)	
XX	DE	Human megakaryocyte stimulating factor (MSF, CACP).	
XX	KM	Human; CACP protein; camptodactyly-arthropathy-coxa vara-pericarditis;	
XX	KW	MSF; megakaryocyte stimulating factor; synovial lubricant;	
XX	KW	chromosome 1q42-31; osteoarthritis; joint lubrication; osteopathic;	
XX	OS	antiarthritic.	
XX	OS	Homo sapiens.	
XX	PN	WO200107068-A1.	
XX	PD	01-FEB-2001.	
XX	PF	21-JUL-2000; 2000WO-US20002.	
XX	PR	23-JUL-1999; 99US-0145328.	
XX	PR	19-JUL-2000; 2000US-0145328.	
XX	PA	(UYCA-) UNIT CASE WESTERN RESERVE.	
XX	PI	Warman ML;	
XX	PT	WPI; 2001-182721/18.	
XX	PT	New composition comprising the camptodactyly-arthropathy-coxa	
XX	PT	vara-pericarditis protein in combination with an anesthetic, useful for	
XX	PT	treating osteoarthritis, or as lubricants of tissue and joints	
XX	PS	Example 1; Page -: 34pp; English.	
XX	XX	The invention relates to a method of treating osteoarthritis via the	
XX	XX	administration of a composition comprising the camptodactyly-arthropathy-	
XX	XX	coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.	
XX	XX	The composition may further comprise a local anesthetic. The composition	
XX	XX	of the invention may be administered via intra-articular or intravenous	
XX	XX	injection. The human CACP protein is identified in the invention as	
XX	XX	being megakaryocyte stimulating factor (MSF). The gene encoding	
XX	XX	CACP protein (MSF) is located on chromosome 1q25-31, and mutations in	
XX	XX	this gene are responsible for the heritable disorder camptodactyly-	
XX	XX	arthropathy-coxa vara-pericarditis, in which patients have synovial	
XX	XX	hyperplasia without evidence of inflammation. CACP protein (MSF)	
XX	XX	acts as a synovium lubricant, and can be used to lubricate tissue and	
XX	XX	joints in the treatment of osteoarthritis. The composition may be	
XX	XX	applied to reduce the symptoms of osteoarthritis (e.g., joint pain,	
XX	XX	loss of range of movement or joint damage). The present sequence	
XX	XX	represents human megakaryocyte stimulating factor (MSF, CACP protein).	
XX	XX	Note: This sequence is not given in its entirety in figure 4 of the	
XX	XX	specification, although a GenBank accession number was given. This	
XX	XX	sequence was therefore obtained from GenBank (U70316).	
XX	XX	Sequence 1404 AA:	

Query Match	99.6%	Score 7245.5	DB 22	Length 1404
Best Local Similarity	97.1%	Pred. No. 0		
Matches 1363	Conservative 0	Mismatches 0	Indels 41	Gaps 1
QY	1	MAWKLPYLLILLISVFIYQOVSSO-----	25	
Db	1	MAWKLPYLLILLISVFIYQVSSGDLSCAGRGEGYSRATCNCdYnqchymccpdf	60	
QY	26	-----ETSCRGCESEFEGRGECDDAOCKKYDKCCPDYSEFCAEVHNPIPSPSKAP	79	
Db	61	KVCTAELackgrcfesfergredcdaqckkydkccpdysefcaevmptpskkap	120	
QY	80	PSGASQITKSTTKSPKPNKKTKVYESEITEHHSVSNQSSSSSSSSSSSTW	139	
Db	121	psgsaqtlkstkspkpnkktkkvlseeliehnsvsnqssssssssstlw	180	
QY	140	KIKSSKNSAANRELOKLLVKNKKNRTKKRPKPVVDEAGSLDNGDKVYTPDST	199	
Db	181	kikssnsaanrelqkllvkndknrtkkpkpvpvdeagsldngdkvtytpdst	240	
QY	200	TOHNKYSTPKITTYAPTPRESIPNSQTSWETLTVKKEITVEKETTNNKQTSNG	259	
Db	241	tgnkvstspklttapyiprpslpsnslsctsltvnketvelcttnkqtsldg	300	
QY	260	KEKTVSAKETQSIEKTSANDLAPTSKVYLAKPYPKAETTTKGALTPPKERTPTKEPAS	319	
Db	301	kekttsakeqsteksakdlaptskvlakpypkaetctkpaltpkertytpkpas	360	
QY	320	TYTKEPPTITKAPPTTKEPAATTTKSAPTTKEPAATTTKEPAATTTKEPAATTTKAP	379	
Db	361	tytkeptptitkapttkepaatttksapttkepaatttkepaatttkepaatttkcp	420	
QY	380	APPTTSAPPTPEPAATTPPKKAPPTTKEPAATTPKEPTPTTTPKAPATTKAPATTKAP	439	
Db	421	appttsapptkepaattpkkappttkepaattpkeptpttppkapaattkapaattpk	480	
QY	440	EPAPITAPKKAAPTTPKEPATTTTKEAPATTTTKRSPPTTKEPATTTKSAPTTKBPAT	499	
Db	481	epapitapkkapittpkepatittpkepatittkrpspttkepatitksapittkbpap	540	
QY	500	TTKSAPPTKEPSPTTKEPAATTPKEPATTTKKBAPTTPKBPATTPKEPATTTKKP	559	
Db	541	ttksapptkepspttkepaattpkepatittkkbapttpkbpattpkapaattpkka	600	
QY	560	APTAPEPATTPKEKELAPTTPKKLTPPTPKKLAPTTPEKAPATTPPELATTPPEBPTPT	619	
Db	601	aptapepatitpkelatittpkkltpptpkklapttpekapattppeelattppebppt	660	
QY	620	PEEPAPPTPKAAAPNTPEKAPATTPKEPATTTKEPATTTKEPATTTKGAATTLKEP	679	
Db	661	peepapitpkaaapntpkapaattpkepatittpkepatittpkepatitpkgaatllke	720	
QY	680	APPTPKKPAPKEIAPTTTKEPTSTTSUKAPATTPPKGAPTTTKEPATTTKEPATTPKG	739	
Db	721	aptpkpkpakeiapittkcptsttsdkpapttkgapttkepatittpkepatitpkg	780	
QY	740	TAPITLKEPATTPPKKPAKELAPTTTNGSTTSDBAPATTPKEIAPTTTKEPATTPPK	799	
Db	781	tapitlkepattpkkpakeiapittkgsttsdbapattpkeiapittpkepatitpk	840	
QY	800	KPAPTPTPEPPTTSEVSTPTTKEPTTIHKSPOBSPBLSAEPPTKALNSPKEPVPPT	859	
Db	841	kpapttpeppttsevspittkcptitihkspsdespbelsaepctkalenspkepvpt	900	
QY	860	TKTPAATPEPNTTAKOKTTERODJRTPEPTTAAPKMTKETATTTKTESKINATTTQV	919	
Db	901	tktpaatpepnttakoktterrjdrtpeptttaaPKMTketatTTkTeskitatTTqv	960	
QY	920	TSTTTQDTPPKITTLKTTTLLAPVYTTTKTITTTTELINKEEELAKRDAATNSKATTPK	979	
Db	961	tsttqdtppkittlkttlapvytttkitttTelnmkpeelakrdaatnskattpk	1020	

```

FT      Region      1266..1331
FT      Region      /Label- Exon_X
FT      Region      1331..1373
FT      Region      /Label- Exon_XI
FT      Region      1373..1404
FT      Region      /Label- Exon_XII
XX      MO9213075-A.
XX      06-AUG-1992.
XX      17-JAN-1992: 92MO-US00433.
XX      18-JAN-1991: 91US-0643502.
XX      10-SEP-1991: 91US-0757022.
XX      (GEMV ) GENETICS INST INC.
XX      Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K;
XX      WPI: 1992-284660/34.
XX      N-PSDB: AAQ27223.
XX      New human mega-karyocyte stimulating factors - for treating
XX      immune deficiencies, cancer, exposure to radiation or drugs,
XX      bacterial and viral infections, etc.
XX      Claim 1, 2 and 3; Fig 1; 87pp; English.
XX      The sequence given is a full length translation from the megakaryocyte
XX      stimulating factor (MSF) precursor. The sequence covered by exons II,
XX      III and IV encodes megakaryocyte stimulating factor (MSF). This
XX      sequence is modified by the addition of an N-terminal sequence encoding
XX      a secretory leader, an initiating methionine preceding exon II and a
XX      terminating codon following exon IV. The cDNA sequence given contains
XX      (meg-CSF). Exon I contains the initiating methionine, and encodes a
XX      classical mammalian protein secretion signal sequence. The sequence
XX      encoding the original meg-CSF includes exons II-IV and is thought to
XX      terminate in the region between amino acid residues 134 - 147. The
XX      primary transcript of this gene may be cleaved in different ways to
XX      yield a family of mRNAs each encoding a different MSF protein. Exons
XX      V and VI are thought to be related to the activity of the factor and
XX      are also implicated in the stability, folding and processing of the
XX      molecule. These exons are also thought to play a role in the observed
XX      synergy of MSF with other cytokines. Exons V - XII are believed to be
XX      implicated in the processing or folding of these exons may contain
XX      the resulting factor, ie. one or more of these exons may contain
XX      sequences which direct proteolytic cleavage, adhesion, organisation of
XX      the cellular matrix or extracellular matrix processing. Both naturally
XX      occurring and non-naturally occurring MSF's may be characterised by
XX      various combinations of alternatively spliced exons from this sequence,
XX      with the exons spliced together in differing orders to form different
XX      members of the MSF family.
XX      Sequence 1404 AA:
XX
XX      Query Match      99.68; Score 7245.5; DB 13; Length 1404;
XX      Best Local Similarity 97.18; Pred. No. 0;
XX      Matches 1363; Conservative 0; Mismatches 0; Indels 41; Gaps 1;
XX
QY      1 MANKTLPYLLLSVFIQOVSSQ-----
DB      1 mawkltpllyllllsvfivqvasqglscagrcgegyrdatcncdyncqhyamecpdf 60
QY      26 -----ELSGKRCFESEFERGECDCDAQCKRYDKCCPDYESFCAEVNNPTSPSSKKAP 79
DB      61 kvtctaelseckgrcfesfergrecdcaqckkydkcpdyefceavmnpicspskxap 120
QY      80 PPSGASQITKSTTKRSPKPPNKKTKTKVSESEITEEHSYSVSNQSSSSSSSSSSSTIIV 139
DB      121 ppsgaqitktsttkrpspkppnkktktkvseelteenhsvensengessssssssstliw 180

```

```

QY      140 KIKSSKNSAANRELOKIKLVKDNKKNRKTKKPTKPPVVDAGSGDNGDFKVTTPDPTST 199
DB      181 kiksksnsaanrelqkllkvdknknrktkkppvvdagsgdngdfkvttpdptst 240
QY      200 TQHNKVSSTSPKITTAKPINPESLPPNSDTSKENSILVANKETTYETKETTNNKQISTDGC 239
DB      241 tqhnkvsstspkittakpinpesslppnsdtskesilvanketttvetkettlnkqistsdcg 300
QY      260 KEKTSARETOSIEKTSKDLAPTSSKVLAKPTPKAETTTKGPALTTTPEPTPTTPKPPAS 319
DB      301 kektsaretsiektsskdlaptsksvjakptkaetttgpaltttpepttpkppas 360
QY      320 TTPKEPTPTTKSAPTTPKBPATTPPKBPATTPPKBPATTPPKBPATTPPKBPATTPPKBP 379
DB      361 ttpkeptpttksapttpkbpattpkbpattpkbpattpkbpattpkbpattpkbpattpkbp 420
QY      380 APTTTSAPTTPKBPATTPPKBPATTPPKBPATTPPKBPATTPPKBPATTPPKBPATTPPK 439
DB      421 aptttsapttpkbpattpkbpattpkbpattpkbpattpkbpattpkbpattpkbpattpk 480
QY      440 EPAPTAKKPPATTPPKBPATTPPKBPATTPPKBPATTPPKBPATTPPKBPATTPPKBPAT 499
DB      481 epaptakppattpkbpattpkbpattpkbpattpkbpattpkbpattpkbpattpkbpattpk 540
QY      500 TTKSAPTTPKBPESPTTKBPATTPPKBPATTPPKBPATTPPKBPATTPPKBPATTPPKBP 559
DB      541 tksapttpkbpespttkbpattpkbpattpkbpattpkbpattpkbpattpkbpattpkbp 600
QY      560 APAPKBPATTPPKBPATTPPKBPATTPPKBPATTPPKBPATTPPKBPATTPPKBPATTPPK 619
DB      601 apapkbattpkbpattpkbpattpkbpattpkbpattpkbpattpkbpattpkbpattpkbp 660
QY      620 PEPAPATTPKKAAPNTPKBPATTPPKBPATTPPKBPATTPPKBPATTPPKBPATTPPKBP 679
DB      661 pepapattpkkaapntpkbpattpkbpattpkbpattpkbpattpkbpattpkbpattpkbp 720
QY      680 APTTPKPPKPELAPTTPKBPATTPPKBPATTPPKBPATTPPKBPATTPPKBPATTPPKBP 739
DB      721 apttpkppkpelattpkbpattpkbpattpkbpattpkbpattpkbpattpkbpattpkbp 780
QY      740 TAPTTKEAPATTPPKBPATTPPKBPATTPPKBPATTPPKBPATTPPKBPATTPPKBPAT 799
DB      781 tapttkeapattpkbpattpkbpattpkbpattpkbpattpkbpattpkbpattpkbpattpk 840
QY      800 KPAPTPEPTPTTSEVSTPTTKPTTIHKSPEBSTPELSAEPTRPALENSPKPEPGVPT 859
DB      841 kpaptpeptpttsevspttkpttihkspebstpelsaeptrpalenspkpepgvpt 900
QY      860 TKTPAATPEMTTAKDRTTERDRLRTTPEYTTAATPKMTKEATTTKETSSTKITTATTTQV 919
DB      901 tktpaatpemttaakdrtterdrlrttpeyttaatpkmtkeatttketsstkitatttqv 960
QY      920 TSTTTODTTPKITTTLKTTTTLAPKVTYTTKATTTTTEIMNRPBETAKPKDRAATNSKATTPK 979
DB      961 tstttodttpkitttlktttllapkvtyttkattttteimnrpbetakpkdraatnskattpk 1020
QY      980 POKPTPAKKKPTSTKPKRTAPRVKPKPTTPTPKMTSTIMELMPTSRIRAEAMLOTTTRPN 1039
DB      1021 pqkptpakkkptstkpkrtaprvkpkpttptpkmtstimelmptsriraeamlootttrpn 1080
QY      1040 QTPNSKLVEVNPKSSEDAAGAGSETPHMLLRPHVFNKEVTPDMQYLPVRVNOGIINPMLS 1099
DB      1081 qtpnsklvevnpkssedagagsetphmlrphvfnkevtpdmqylprvpngiilnplms 1140
QY      1100 DETNINCKRPVDLITLLRNQTLVAFRGHYFWMISPPSPSPSAPARTTEWGIJSPIDVFT 1159
DB      1141 detninckrpvdlltllrnqtlvafrghyfwmisppspspsaparttewgijspidvft 1200
QY      1160 RCMCEKTEPFKHSQSWRTTNDIKDAGYRPPIFKGFGTLGQIVAAALSTAKYKKNPESVY 1219
DB      1201 rcncegkttfifkdsqywrftndikdagyrpifikfgglqglvaalstakykknpeasy 1260

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM protein - protein search, using SW model

Run on: April 26, 2002, 16:24:09 ; Search time 138.84 Seconds
(without alignments)
727.182 Million cell updates/sec

Title:	AA2
Perfect score:	7276
Sequence:	1 MAWKLPYLLLLSVFVIO.....ARAITRSGQTLKVMYNCP 136

Scoring table: `pbccore2`
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB	seq length:	0
Maximum DB	seq length:	20000000000

Post-processing:	Minimum	Match	0%
	Maximum	Match	100%
	Listing	first	45 summaries

1:	/S1D58/gcgdata1/geneseq/geneseqp/AA1960.DAT *
2:	/S1D58/gcgdata1/geneseq/geneseqp/AA1961.DAT *
3:	/S1D58/gcgdata1/geneseq/geneseqp/AA1962.DAT *
4:	/S1D58/gcgdata1/geneseq/geneseqp/AA1963.DAT *
5:	/S1D58/gcgdata1/geneseq/geneseqp/AA1964.DAT *
6:	/S1D58/gcgdata1/geneseq/geneseqp/AA1965.DAT *
7:	/S1D58/gcgdata1/geneseq/geneseqp/AA1966.DAT *
8:	/S1D58/gcgdata1/geneseq/geneseqp/AA1967.DAT *
9:	/S1D58/gcgdata1/geneseq/geneseqp/AA1968.DAT *
10:	/S1D58/gcgdata1/geneseq/geneseqp/AA1969.DAT *
11:	/S1D58/gcgdata1/geneseq/geneseqp/AA1970.DAT *
12:	/S1D58/gcgdata1/geneseq/geneseqp/AA1992.DAT *
13:	/S1D58/gcgdata1/geneseq/geneseqp/AA1993.DAT *
14:	/S1D58/gcgdata1/geneseq/geneseqp/AA1994.DAT *
15:	/S1D58/gcgdata1/geneseq/geneseqp/AA1995.DAT *
16:	/S1D58/gcgdata1/geneseq/geneseqp/AA1996.DAT *
17:	/S1D58/gcgdata1/geneseq/geneseqp/AA1997.DAT *
18:	/S1D58/gcgdata1/geneseq/geneseqp/AA1998.DAT *
19:	/S1D58/gcgdata1/geneseq/geneseqp/AA1999.DAT *
20:	/S1D58/gcgdata1/geneseq/geneseqp/AA2000.DAT *
21:	/S1D58/gcgdata1/geneseq/geneseqp/AA2001.DAT *
22:	/S1D58/gcgdata1/geneseq/geneseqp/AA2001.DAT *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB	ID	Description
1	7245.5	99.6	1404	13	AAR26049	MSF precursor. Sy
2	7245.5	99.6	1404	22	AAB60568	Human megakaryocy
3	7245.5	99.6	1404	22	AAB29773	Human megakaryocy
4	6669.5	91.7	1289	22	AAB29772	Human EST encoded
5	3484	47.9	902	22	AAR29788	Human MSF-derived
6	1916.5	26.3	452	16	AAR80041	Human megakaryocy
7	1707.5	23.5	472	22	AAB60569	Bovine MSF ortholo
8	1175	16.1	5179	22	AAM24516	CG999 predicted an
9	981	13.5	763	21	AAG38942	Arabidopsis thalian
10	950	13.1	1664	19	AAW43106	C. thermophilum O
11	805	11.1	1325	22	AAW03645	Peptide #2327 encc

12	715	8.8	763	18	AAV31852
13	645.5	8.9	4412	21	AAV53666
14	612	8.4	572	18	AAV31855
15	555.5	7.6	844	7	AAV60570
16	542	7.4	807	21	AAV54467
17	520.5	7.2	788	21	AAV54466
18	513	7.1	1837	21	AAV17276
19	507	7.0	182	9	AAV10872
20	506.5	7.0	744	9	AAV82975
21	498.5	6.9	1721	21	AAV11727
22	498.5	6.8	1721	19	AAV48299
23	496.5	6.8	2971	21	AAV41231
24	489.5	6.7	3272	22	AAV50363
25	489.5	6.7	3118	22	AAV50362
26	488.5	6.7	826	13	AAV26042
27	488	6.7	617	22	AAV16458
28	488	6.7	617	22	AAV04187
29	488	6.7	957	21	AAV59248
30	488	6.7	957	22	AAV24513
31	485	6.7	1127	22	AAV95541
32	476.5	6.5	652	9	AAV22974
33	467.5	6.4	511	22	AAV14883
34	467.5	6.4	511	22	AAV027312
35	467.5	6.4	511	22	AAV02667
36	454.5	6.2	1012	22	AAV174048
37	454	6.2	2819	12	AAV35408
38	450.5	6.2	378	22	AAV4160
39	448	6.2	1237	21	AAV81691
40	447.5	6.2	3266	21	AAV84269
41	446.5	6.1	378	22	AAV4162
42	446.5	6.1	750	20	AAV05477
43	446.5	6.1	2665	22	AAV45533
44	446.5	6.1	2665	22	AAV69590
45	446.5	6.1	2665	22	AAV02255

ALIGNMENTS

Mycobacterium tuberculosis
 sequence g1101742
 Mycobacterium tuberculosis
 Sequence of the Fa
 Amino acid sequenc
 Amino acid sequenc
 Cryptosporidium pa
 Protein encoded by
 Bioadhesive precu
 Portion of Cryptos
 Cryptosporidium pa
 Human ORF4 ORF995
 Human SRAP. Homo
 Human SRAP. Homo
 P. yoelii Ssp2 ant
 Peptide #2892 enco
 Peptide #2869 enco
 Human MUC11 polype
 C9orf9 predicted am
 Human protein sequ
 Bioadhesive precu
 Peptide #1317 enco
 Peptide #1349 enco
 Peptide #1289 enco
 Human atropin-1 r
 Human OTCG57 gene
 PRP 378. Triflicum
 Streptococcus pneu
 Human ORFX ORF225
 PRP encoded by chi
 C. albicans Rbl1 F
 Peptide #967 enco
 Peptide #997 enco
 Peptide #941 enco

1-3

Db 811 epsttge--hadekkasegdnislrseeteek 842

Search completed: April 26, 2002, 16:24:52
Job time: 482 sec

Search completed: April 26, 2002, 16:24:52
Job time: 482 sec

```

OY 919 -RKMTSTPELNPTSRIAMLOTTT 943
      : : | | | | | | | | : :
Db 4258 iKdIlIteSeIvgsaIfeclvpspt 4283

RESULT 14
AAW31855
ID AAW31855 standard; Protein; 572 AA.
XX
AC AAW31855;
XX
DT 27-APR-1998 (first entry)
XX
DE Mycobacterium tuberculosis 55 kDa protein.
XX
KM Tuberculosis; mycobacteria; infection; diagnosis;
XX antimycobacterial; antibiotic; vaccine.
XX
OS Mycobacterium tuberculosis.
XX
PN MO9741252-A2.
XX
PD 06-NOV-1997.
XX
PF 18-APR-1997; 97WO-EP01973.
XX
PR 29-APR-1996; 96DE-4017184.
XX
PA (GBFB ) GBF GES BIOTECH FORSCHUNG GMBH.
XX
PI BspItia C, Honisch C, Moreno C, Singh M;
XX
DR WPI; 1997-549750/50.
XX
DR N-PSDB; AAT93610.
XX
PT New DNA and related proteins or RNA derived from M. tuberculosis
PT used for diagnosis of mycobacterial infections, monitoring
PT vaccination and development of anti-mycobacterial agents
XX
PS Claim 11; Fig 16; 55pp; English.
XX
CC This novel 55 kDa protein is encoded by an open reading frame of
CC a Mycobacterium tuberculosis DNA fragment (see AAT93610) containing
CC polymorphic GC-rich sequences. Its amino acid sequence shows
CC a high proline content, but there is no homology to any known
CC proline-rich antigens of mycobacteria. Novel M. tuberculosis
CC proteins (see AAW31851-57) are claimed. These can be produced as
CC recombinant proteins, especially in bacterial, yeast, fungal or
CC higher eukaryote host cells, and used for diagnosing tuberculosis
CC and other mycobacterial infections in humans or animals. The
CC claimed proteins can also be used for epidemiological studies, for
CC monitoring vaccination, and for the development of vaccines and
CC anti-mycobacterial drugs.
XX
SO Sequence 572 AA;

Query Match 9.0%; Score 612; DB 18; Length 572;
Best Local Similarity 30.6%; Pred. No. 5.5e-29;
Matches 193; Conservative 29; Mismatches 253; Indels 156; Gaps 33;

```

```

OY 361 -----PKRPA-----PTTPKEPAPTTTKEPSPTTKEPAPTTTKSAPTTTKEPAPTT 407
      | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 166 evIaGaiIpsrIpsrIltttIpaIpaI-----pIppIppIppIIntaIpaI-----pIppIpp 217
OY 408 TKSAPTTPK-EPSPTTTPKEP--APTTPKEPAPTTTTPKKAPTTTTPKEPAPTTTKEPAPTTTK 464
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 218 taIappIppIaIppIspgyvpaIpppIpppIpppIppIppIppIppIppIppIppIppIppIpp 276
OY 465 KPAFTAAPKPPAPTTTKEPAPTTTTPKKIppTTPPE--KIAPTTPKEPAPTTTPEELAPTTPEEP 522
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 277 knppap--peppekspsaIpppapppsaavrvppspIppap-----paapras 327
OY 523 TPTEEPAPTTTPKA---AAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTPK---ETAPTTPK 576
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 328 mpaIpp--pappspatIppIppIppspapusp--pappap-----pIppkIIsaIppcpIppv 379
OY 577 GTAFTTKEPAPTTTPKKAPKELAPTTTKEPSTSDKAPTTTPKGAFTTTPKEPAPTTTP 636
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 380 ppaIpp--rppapapap--pappel-----pappdp-----pIppvausapap 416
OY 637 KEPAFTTTPKGTAFTTTPKEPAPTTTPKKAPKELAPTTTKGP-----TSTSDKAPATTTP 689
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 417 --pappapsaIpp--IvnpappIppaIppaIppaIppaIppaIppaIppaIppaIppaIppaIpp 473
OY 690 KEFA--PTTPKEPAP--TTTPKKAPTTTPEPTTSEVSTPTTKEPTTIKSPDESTP 744
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 474 nsmaIpppappdppIppIppIpppappIppmsppappIppaIppaIppaIppIppIppIppIpp 530
OY 745 ELSAEPTPKALENSPKKEGVTTKTPAATKP 775
      | : | | | | | | | | | | | | | | | | | | | | | |
Db 531 pla-----pvygaIppIppIppIpp 549

RESULT 15
AAP60570
ID AAP60570 standard; Protein; 844 AA.
XX
AC AAP60570;
XX
DT 24-AUG-1991 (first entry)
XX
DE Sequence of the Falciparum Interspersed Repeat Antigen
DE (FIRA).
XX
KM Malaria vaccine; antigen; epitope.
XX
OS Plasmodium falciparum.
XX
PN WC8601802-A.
XX
PD 27-MAR-1986.
XX
PF 11-SEP-1985; 85WO-0006960.
XX
PR 11-SEP-1984; 84AU-0007067.
PR 11-SEP-1984; 84AU-0007066.
PR 10-SEP-1985; 85AU-0047326.
XX
PA (HALL-) HALL INST MED RES.
XX
PI Kemp DJ, Anders R, Coppel RL, Brown G, Saint RB, Cowman AF;
XX
XX WPI; 1986-094065/14.
XX
XX N-PSDB; AAN60473.
XX
PT DNA coding for Plasmodium falciparum antigens - expressing
PT poly(peptide(s) having antigenicity of RESA or FIRA antigens of P
PT falciparum
XX
PS Disclosure; Fig 7; 55pp; English.
XX
CC The inventors claim a novel DNA molecule which comprises a
CC nucleotide sequence corresp. to all or a portion of the base

```

```

Db      638 paapapapvratcp-----ppapappppn 665
Oy      880 SKATPRPKP-----TKAKKPTSTIKKPKTMRVKKPKTTPPRKMTSTM-----PELNP 930
Db      666 smalppppdpplplacppappplpmppapplppappppppllnqpppplap 725
Oy      931 TSRIAMMQTTTRPNQTRNSKIVEVNPKSEDAAGEGET 970
Db      726 vpgaplaplplngprvafarknsl-----gssgac 756

RESULT 13
AAV53666
ID      AAV53666 standard; Protein: 4412 AA.
XX
AC      AAV53666;
XX
DT      22-FEB-2000 (first entry)
XX
DE      Sequence g1/1017427/emb/CAA62189 from an alignment with protein 608.
XX
KW      Mechanical stress; gene therapy; protein 608; osteoporosis; bone density;
XX      bone development; g1/1017427/emb/CAA62189.
XX
OS      Unidentified.
XX
PN      WO9960164-A1.
XX
PD      25-NOV-1999.
XX
PE      14-MAY-1999; 99WO-US11066.
XX
PR      15-MAY-1998; 98US-0085673.
XX
PA      (OVAR-) OVAR BIOTECH INC.
XX
PI      Elhat P, Mor O, Skalter R, Feinstein E, Faerman A;
XX      WPI: 2000-053304/04.
XX
PT      Identification of stress induced genes for determining risk and
XX      preventing, treating or controlling osteoporosis
XX
PS      Claim 32; Fig 6A-R; 308pp; English.
XX
CC      The present sequence is obtained from a clustral X alignment with
CC      protein 608. Protein 608 was identified using the method of the invention
CC      after subjecting rat osteoblasts to mechanical stress. Expression of the
CC      608 gene was found to be upregulated by about 3-fold in cells subjected
CC      to mechanical strain. The specification describes a method for the
CC      identification of genes responsive to a specific mechanical stress. The
CC      method comprises applying the mechanical stress to an organism (tissue
CC      or cells comprising bone cells), isolating the specifically cellular
CC      fractions and extracting mRNA from them, and differentially analysing the
CC      mRNA in comparison with control samples. The method is used to identify
CC      genes whose expression is responsive to a specific stress. The identified
CC      genes are employed in determining risk associated with a physiological or
CC      disease state. The risk determination methods are used for testing a
CC      medicament for gene therapy. These medicaments, or genes identified by
CC      the method of the invention, are used for treating, preventing or
CC      controlling a physiological or disease state (especially osteoporosis or
CC      bone density or other factors causing or contributing to osteoporosis or
CC      its symptoms or other conditions involved in mechanical stress or its
CC      lack. The methods can also be used for advancing research or studies in
CC      bone development.
XX
SQ      Sequence 4412 AA:

```

```

Query Match      9.28; Score 627; DB 21; Length 4412;
Best Local Similarity 24.18; Pred. No. 6.9e-29;
Matches 266; Conservative 110; Mismatches 442; Indels 288; Gaps 51;

```

```

Oy      26 EISCGRCFES-----FERGRCDDAQCCKVKCCPDYESFCA--EVKDNKK----- 71
Db      3278 qlschvgseprlqlwlagre-----lkpdcresfsaigtavlelrdvakedsdyv 3331
Oy      72 -----NRTKKKPTPK-PPVDEA--GSGLDNGDKVYTPDSTTQHKNVSTSPKIT 119
Db      3332 cksanvgsdtkskvltkdkpavapakkaavdgrlffvsepsqlrvvektatflakv 3391
Oy      120 TAKPINRPSLPPNSDTSKETSLELVNK-----ETTVETKETTTN----- 159
Db      3392 gqddp-----pnvkwtkgvrqlnggvrvfingqdeaklelrdtclsdgjlrcvaf 3444
Oy      160 -----KQSTDGKEKTTSAKETOSIEKTSIA-----KDLAPT 190
Db      3445 nehgeiesnvlgyderkkgekieg-dlramlkktrpllkkgaageeeldmellknvdpk 3503
Oy      191 S-----KVLAPTPKAEETTTGPA 209
Db      3504 eyekyarmygltdfrglqrveeehrevkhrvlefeaevevfekp---kappkgye 3559
Oy      210 LT-----TPKEPTPTTP-KEPASTTPKEPTPTTIKSAPTTPKEP-APTTPKSAPTTPKEP 262
Db      3560 lsekliipkkpctkvprkeppakvpevpkivveekvrypeervpctkvpevlppkev 3619
Oy      263 APTTTKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKK---PAPTTPKEPA 318
Db      3620 vp-----ekkvvpvppakkpeap-----pkkvpeapkevppekkvvpvppkkpevp 3665
Oy      319 PTPPEPTPTTPKEPAPTTPKEPAPTTP-----KEPAPTAPKKPAPTTPKEPAPTTPKE 371
Db      3666 tkvpevpkaavpekkvpealipkpespppevfespeespaapx-kvevpvrivevpke 3724
Oy      372 PAPTTKEPSEPTTPKEPAPTTPKSAPTTKEPAPTTPKSAPTTKEPAPTTPKEPAPTTP 431
Db      3725 vvpkekkyvpaap--pkpdevpvkv-vpeapkevppekkvvpv--ppkkvpevpctk-----vp 3775
Oy      432 KEPAPTTPPKRAPPTTPKEPAPTTPPKRAPPTTPKKRAPATAKEPAPTTPKETAPTTPKKLT 491
Db      3776 evpkvavpekkvp-----ealppkpespppevfespeesvalpepaeveeepaapqvt 3831
Oy      492 ----PTTPEKLAPTT---PEKPAPTTPBELAPTTPPEPTTP-PTTPEAPTTPKAAPNT 542
Db      3832 vvpknvpekkapavakpkelppvkvppevpkevppekkvplvvpkk--peapvakvpev 3889
Oy      543 PKEPAP-----TTPKEP-----APTTPKEPAPTTPKETAPTTPKGAPTTLKEPAPTTPK 592
Db      3890 pkevvppekkvavppkkpevpvakvpevpkpv-----leekpavp---vperaeespppevy 3942
Oy      593 KPAPKELAPTTTKEPTSTSDKAPAPTTPKGTAPTTPKEPAPTTPPKEPAPTTPKGAPTTTL 652
Db      3943 e--peelap---eeelapeekvpvvaee-----eepevpvpaapeekkllipkkvp--vL 3992
Oy      653 KEPAPTTPKKPAPKELAPTTTGGTSTTSDKAPAPTTPKETA-----PTTPKEPAPTTP 705
Db      3993 kkpbeapppkepekev---lekplkprppppapppkedvkekifqlkalipkkvpenp 4049
Oy      706 KKP--APTTPETPTTPPTTSEVS--TPTTTPKEP-----TTIHKSPEDSPTPELSAETTPVA 754
Db      4050 qvpekelclpkkvggekkvkrkllperkpepeevnlksvlrtfpeeeekve---pkk 4105
Oy      755 LENSPEKPGVPTTTPAATKP-----EMTTAKDKTTERDLRTPTTAAPKKTKETAATV 810
Db      4106 le-kvkkpavp---eppppkvveevpvtckrekliprpkvkelkpalpjpapepkkk 4161
Oy      811 TEKTTESKITATTTQVSTTTQDTTTPFKITTLTKTTTLAKRVTTTKKTTTTEIMNK---- 866
Db      4162 pea-----evkltikppvepepilaapvlpvvgvkkkaeak 4197
Oy      867 --PEETAKPKDRATNSKATTPKPKPTKAPKKPTS--TKKPTTMPVRVKPTTPP----- 918
Db      4198 apkeeaekpgrlkyppkktppieaerkltpgsgeekpdeapltvqlkavpklfveke 4257

```

```

OY 532 PTPKAAAPNT---PKPAPPTTPKEPAPPTTPKAPPTTPKGAAPTTPKLEPAP 588
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
PS 634 enrerlanekltsspaepengrtfpfanekltsspaephegertplanentlspaep 693
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 589 -----TPKKRPAKELAPPTTKEPTSTSDKRAPPTPKGTATPTPKPAPPT 642
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 694 tenrerlanektfpaepenrerlanentcspqplengrtfpfanekltspaept 753
OY 643 -----TPKGAAPTTPKLEPAPT-----TPKKRPA-PKE-----LAP 670
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 754 engkrfpfanekltsspaephegertplanentlsspaepenrerlanekltqfpaep 813
OY 671 TTTKGPTSTSDKRAPPTPKET-----APTTPKRAPPTTPKRAPPTTPPTPT 749
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 814 tenrerlanektfpaepenrerlanentlsspaephegertplanekltspaept 873
OY 720 TSEVSTPTTPKEPTTIHKSPE-----STPELSAEPPTPK-----ALENSPKPEGV 764
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 874 engertfpfanekltsspaephegertplanentlsspaephegertlanekltspak 932
OY 765 PT-----TKPAAATKP-----EMTTAKDKTTERDLTTPETTTAAFKMT 804
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 933 ptehgertvnedtsspaepengertplanentlsspaephegertlanekltspaept 992
OY 805 KE-----TATTEKTESKITAT-TGYTSSTTODTTPFKI-----TTL----- 842
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 993 tengerfpfanekltspakpthehemtspantcspvkvkhegertlanekltsp 1052
OY 843 -----KTTTLAPKVTYT-----KTTTTEIMNKPEETAKPKDRATNSKATP 885
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1053 egrtpehgaaktssanekltspakpthehemtspantcspvkvkhegertlanekltsp 1112
OY 886 KPQKP-----TKAKRKPSTSKKPTMPVRKP----- 912
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1113 apaeplhakrtllahakmtqvtleakstehepeltste-Ktlrtpekllysektictg 1171
OY 913 KTPPTPKMTSTMPELNPTSRIAEAMLTQTTTRPNOT 948
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1172 kntrvpekpentcgtlcttcltlakvksstenpekt 1207

RESULT 12
ID AAW31852 standard; Protein: 763 AA.
XX AAW31852:
XX
AC AAW31852:
XX
DT 27-APR-1998 (first entry)
XX
DE Mycobacterium tuberculosis 74 kDa protein.
XX
KW Tuberculosis; mycobacteria; infection; diagnosis;
KW antimycobacterial; antibiotic; vaccine.
XX
OS Mycobacterium tuberculosis.
XX
PN W09741252-A2.
XX
PD 06-NOV-1997.
XX
PE 18-APR-1997; 97WO-EP01973.
XX
PR 29-APR-1996; 96DE-4017184.
XX
PI Espitia C, Honisch C, Moreno C, Singh M;
XX (GBFB) GBF GES BIOTECH FORSCHUNG GMBH.
XX
PI
XX WPI: 1997-549750/50.
XX
DR N-PSDB: AAT93610.
XX
PT New DNA and related proteins or RNA derived from M. tuberculosis -
used for diagnosis of mycobacterial infections, monitoring

```

```

PT vaccination and development of anti-mycobacterial agents
XX
XX Claim 5; Fig 13; 55pp; English.
PS
XX
XX This novel 74 kDa protein is encoded by an open reading frame of
CC a mycobacterium tuberculosis DNA fragment (see AAT93610) containing
CC polymorphic GC-rich sequences. Its amino acid sequence shows
CC a high proline content, but there is no homology to any known
CC proline-rich antigens of mycobacteria. Novel M. tuberculosis
CC proteins (see AAW31851-57) are claimed. These can be produced as
CC recombinant proteins, especially in bacterial, yeast, fungal or
CC higher eukaryote host cells, and used for diagnosing tuberculosis
CC and other mycobacterial infections in humans or animals. The
CC claimed proteins can also be used for epidemiological studies, for
CC monitoring vaccination, and for the development of vaccines and
CC anti-mycobacterial drugs.
XX
XX Sequence 763 AA:
SQ

Query Match 10.5%; Score 715; DB 18; Length 763;
Best Local Similarity 28.2%; Pred. No. 5, 2e-35;
Matches 248; Conservative 39; Mismatches 361; Indels 232; Gaps 39;

OY 197 PTPKAEETTKGPAALTTPKEPTTPTPKKEPASTTPKEPTPTTIKSAPTTPKEPATTKSAP 256
DB 3 pvp-----apalalaplppappapaepekpfppap-----pappcwmvlvsap 46
OY 257 TTPKEPAPPTTPKEPAPPTTPKEPAPPTTTSAPPTTPKEPAPPTTPKRAPPTTPKE 316
DB 47 pcp--pap-----papkpxskapfpvpppapaparelalpp--pap-----peapre 90
OY 317 PAPTTPKEPTPTT-----PKPAPPTTPKEPAPPTTPK-EPAPTAAPK-KPAPTTPKEPAPPTPK 370
DB 91 srpalppcopprrvvlppdppeaaprvppapnspfpfpfpapkfvpapvvp--pvpnspp 148
OY 371 EP-APTTPKEPSTTPKEPAPPTTTSAPPTTPKEPAPPTTTSAPPTTPKEPSTTPKEPAP 428
DB 149 ffpfpalalnpap-----kaplanspplppapfpagf--ppaaprvppapapkskpa 201
OY 429 TTPKEPAPPTTPKRAPPTTPKEPAPPTTPKEPAP-----TTPKKRAPATKAPAPPTTPKETAP 484
DB 202 spptrppap-----pmpatpmeffpplppvppdpjsketppapppappplppavpplppvplp 256
OY 485 TTPKKLTPPTPEKLAP-----TTPKEPAPPTPEELAPPTPEEPPTTPPEEPAPPT-- 534
DB 257 pvpnkilppap--appvavaavlvapocplpplgnhnpapapavpvpvplaplgnshp 313
OY 535 ---PKAAPNTPKKEPAP----- 548
DB 314 pappsapvppvplaplplsgfrpvasvkwgsftlctfcccvcsgevlagalnpsrpspl 373
OY 549 --TTPKEPAPPTTPKEPAPPTTPKETA---PTTPKGA---PTTLKEPAPPTP-KKAPKE 598
DB 374 tttcpalpaplpplpplpplntavvplpplpvltalappllpplplpispvppap-- 431
OY 599 LAPTTTKEPTSTSDKRAPPTPKGTAPPTPKETA-----PTTPKEPAPPTTPKGTATTTL 652
DB 432 --pippgkpwltlpplapapepk--lvpylppgscppsekkpnpapapepeksspalpp 488
OY 653 KEPAATTP---KKPAPKELAPTTTKGP--TSTTSOKPAPTTPKET---APTTPKEPAPPTTP 705
DB 489 appapmpsavrvppspplppappaaprasmpalppapppspalclcpplppppapansp 548
OY 706 KKPAPTTPPEPTTPSEVSTPTTKEPTTIHKSPT---DESTPELSAEPPTPKA--LENSP 759
DB 549 --papppappklls--anppcpvppapnrrpappapapelpappdpdpfpvpaansp 604
OY 760 KEGVPTTKTPAATKEMTTAKDKTTERDLRTTPETTTAAPKMTKEATTEKTESKI 819
DB 605 --papppapppsalplfvnpa-----pfpcaapx-----srpal 637
OY 820 TATTTQVSTTODTTPPKITTLKTTTLAPKVTYTKKTTTTEIMNKPEETAKPKDRATN 879

```


Best Local Similarity 38.5%; Pred. No. 6.3e-51;
Matches 226; Conservative 24; Mismatches 293; Indels 44; Gaps 11;

```

OY 202 ETTTKGPAALTTPEKPPPTTTPKEPASTTPKEPTTTIKSAPTTTKEPAPTITTKSAPTTTKE
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 126 qtdagdrpylrrppprrlppppccvctcp-epsp-----pppplvt---pqilpct 170
OY 262 PATTTKKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 171 lrppllpctppcpptapctevlptcppllpqpllppllppllppllppllppllppllppl
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 320 TTPKEPPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKE
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 227 ntppetpctppctppctppctppctppctppctppctppctppctppctppctppctppct
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 379 EPSPPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 287 etspntppctspctppctppctppctppctppctppctppctppctppctppctppctppct
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 439 PKRPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKE
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 341 ppqplpntppctppctppctppctppctppctppctppctppctppctppctppctppct
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 499 APPTPEKAPPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 401 lrpplpplpntppctppctppctppctppctppctppctppctppctppctppctppct
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 559 PKRPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKE
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 461 pkcp-rltppctppctppctppctppctppctppctppctppctppctppctppctppct
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 619 TPKGTAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKE
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 515 tppqtpnqspqilppllpntppctppctppctppctppctppctppctppctppctppct
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 679 TTSKRP-----APTTPKETAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKE
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 575 lrpntppqilppllpntppctppctppctppctppctppctppctppctppctppctppct
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 732 P---TTTHKSPDESTPELSAEPPTKALENSPKKEGVPPTTKPAATKP 775
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 635 pppttlppspstpspnpspppksqpppprrfcpqpppprgtclp 681

```

RESULT 10
AAW43106
ID AAW43106 standard; Protein; 1664 AA.

XX AC AAW43106;

XX DT 16-OCT-1998 (first entry)

XX DE C. thermocellum OlpB protein.

XX KM Multimer; enzyme; complex; protein-protein interaction; dockerin domain;
cellulose domain; catalytic subunit; scaffold subunit; SdA; synergistic;
cellulosome integrating protein; scaffoldin dockerin binding protein.

XX KM Clostridium thermocellum.

XX XX Key Location/Qualifiers

FT FT Domain 28..192 /note= "cohesin type II domain"

FT FT Domain 207..363 /note= "cohesin type II domain"

FT FT Domain 409..565 /note= "cohesin type II domain"

FT FT Domain 607..763 /note= "cohesin type II domain"

XX XX FR2748479-A1.

XX XX PD 14-NOV-1997.

XX PF 10-MAY-1996; 96FR-0005854.
XX PR 10-MAY-1996; 96FR-0005854.
XX PA (INSP) INST PASTEUR.
XX PI Beguin P, Leibovitz E;
XX DR WPI: 1998-011569/02.
XX DR N-PSDB; AAT86623.
XX PT Cellulase proteins with cohesin or dockerin type II domains - useful
PT for potentiating the activity of multiprotein enzyme complexes
PS Claim 7; Page 31-39; 60pp; French.

CC Multimeric protein, especially enzymatic, complexes are held together
CC by protein-protein interactions between domains designated dockerins
CC and cohesins, which are found on the catalytic and scaffold subunits
CC respectively. An example of such a complex is the cellulose degrading
CC protein complex from Clostridium thermocellum, known as the cellulosome.
CC This complex comprises around 15 proteins including endoglucanases,
CC cellobiohydrolases, hemocellulases, e.g. xylanases or lichenases, which
CC interact with a central "scaffold" protein designated the cellulosome
CC integrating protein (CipA; see AAW43108). The catalytic subunits
CC interact with the Cip subunit via conserved 23 amino acid dockerin
CC domains. Cip has been shown to contain 9 copies of a cohesin domain.
CC The invention relates to the isolation of proteins binding to a novel
CC dockerin type domain found in the C-terminal portion of Cip. The new
CC domain is designated a type II dockerin domain (as compared to the type
CC I domain found on the catalytic subunits of the cellulosome). The type
CC II dockerin domain has some sequence similarity to the type I dockerins
CC but is unable to bind type I cohesin domains.
CC The sequence presented here is an example of a protein which binds
CC the novel type II dockerin domain and is the product of the OlpB gene.
CC The protein contains 4 type II cohesin domains in the N-terminal portion
CC of which the first domain (amino acid residues 28-192) is thought to
CC bind CipA. The novel type II dockerin and cohesin domains can be used
CC in complexes, especially enzyme complexes, to potentiate their catalytic
CC actions in a synergistic manner.

XX XX Sequence 1664 AA;

Query Match 13.9%; Score 950; DB 19; Length 1664;
Best Local Similarity 31.0%; Pred. No. 1.1e-48;
Matches 312; Conservative 104; Mismatches 361; Indels 230; Gaps 55;

```

OY 193 VLAQTP- KAEETTKGPAALTTPEKPPPTTTPKEPASTTPKEPTTTIKSAPTTTKEPAPT
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 758 vviqaprikaasdepilpdlpsdepts-----depts---deptsdeptsd 804
OY 252 TKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKE
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 805 eptpseupdepilpdlpsdeptsdeptsdeptsdeptsdeptsdeptsdeptsdeptsd 864
OY 308 KPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKE
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 865 sdeptsdeptsdeptsdeptsdeptsdeptsdeptsdeptsdeptsdeptsdeptsd 919
OY 367 TTPKEPA--PTTTPKEPST-TPKEPAPTTPKS-APTTPKEPAPTTPKEPAPTTPKEPAPT
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 920 sdeptsdeptsdeptsdeptsdeptsdeptsdeptsdeptsdeptsdeptsdeptsd 979
OY 420 -----PTTTPKEPAPT-TPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 960 psdeptsdeptsdeptsdeptsdeptsdeptsdeptsdeptsdeptsdeptsdeptsd 1039
OY 475 APT---TPKEPAPTTPKRLTTPPEKLAAPTTPKEPAPT---TPKEPAPTTPKEPAPTTP
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1040 tpsdeptsdeptsdeptsdeptsdeptsdeptsdeptsdeptsdeptsdeptsd 1094

```

PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140655.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140823.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 02-JUL-1999; 99US-0142154.
 PR 06-JUL-1999; 99US-0142055.
 PR 08-JUL-1999; 99US-0142380.
 PR 09-JUL-1999; 99US-0142803.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 21-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 03-AUG-1999; 99US-0146389.
 PR 04-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.

PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149375.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150565.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159328.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 22-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 28-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match

14.4%; Score 981; DB 21; Length 763;

(e.g. by enzyme linked immunosorbant assay (ELISA)). AA128460 to AA129512 CC and AA24494 to AA24523 represent nucleotide and amino acid sequences CC given in the exemplification of the present invention.

Sequence 5179 AA:

Query Match 17.2%; Score 1171; DB 22; Length 5179;
Best Local Similarity 32.8%; Pred. No. 2.3e-61;
Matches 364; Conservative 57; Mismatches 461; Indels 228; Gaps 41;

38 EGRGRCDDAO---CKRYDK-----CCPDYE--SFCAEVKDKKKKKKKPPKPPV 84
Db 1350 qhgqkvqcdsvsfglcknedqfgnqpfgydklrvncwpm-----kcltspst 1403
QY 85 VDEAGSGLDNGDFKVTTPDTSTQHNKVSSTPKITAKINRPSLPKSDSKE--TSL 142
Db 1404 ttp-----spptltt-----tlpptspptlttttpppttppptt 1445
QY 143 TVNKEITVETK--ETTTNKKQSTDGKEKETSASKEIOSIEKTSANDLAPTSVLAKPPPK 200
Db 1446 ttppttspstlttpppttsppttspptt-----pspp 1487
QY 201 AETTTGPAITTPKEP--TPPTKEPASTTPKEPTTITKSAPTTPKEPAPTTPKSAPT 258
Db 1488 ttttpppttsppttppmtltp--pactltlpt--tupspttlttpppttspptt 1544
QY 259 PKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKPPAPTTPKEBA 318
Db 1545 tpttppstlttpppttsppttlttpppttspptt-----tppptlttppp 1598
QY 319 PTPPKPPTTPKEPAPTTPKEPAPTTPKEP-----APTAKPPAPTTPKEPAPTTPKE 371
Db 1599 ttpstpttlttpppttsppttsppttlttppstltlpppttpppttpppttpp 1658
QY 372 PAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAP--TPPKESPPTTPKEPAP 428
Db 1659 ptttppstlttpppttsppttspstlttppstlttppstlttppstlttppst 1717
QY 429 TTPKEPAPTTPKPPAPTTPKEPAPTTPKEPAPTTPKPPAPTTPKPPAPTTPKPPAPTTP 483
Db 1718 ttpstppstlttpppttsppttsppttlttppstlttppstlttppstltt 1777
QY 484 PTPP-----KRLTP----- 492
Db 1778 ptpcpcvplcmwtgldsqkpnfhkpggdtelldqvcgpgwaanlsctaanypdvblgq 1837
QY 493 -----TPPEKIAPTTPKEPAPT 510
Db 1838 qtvvcadvsglcknedqkpgvlpmafclynelnvqcecvqptumtltteaptp 1897
QY 511 PEELAPTPPEEPPT-TPPEEAPPTPKAAANPMPKEAPPTTPKEP-----APTTPKEPAP 564
Db 1898 tpttlttvtprpctpgtqtpptlttlttvtprpctpgtqtpptlttlttvtpr 1957
QY 565 T-TPKEAPPTPKGTAPTTLKEPAPTTPKPPAPKELAPTTTKEPTSTSDKPAP--TPPKG 622
Db 1958 tptgtqtppttlttvtprpctpgtqtpptlttlttvtprpctpgtqtppt 2017
QY 623 TAPTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKPPAPKELAPTTTKEPTSTSD 682
Db 2018 tttvtprpctpgtqtp--tptlttlttvtprpctpgtqtpptlttlttvtprp 2076
QY 683 KPAP--TPPEKIAPTTPKEPAPTTPKPPAPTTPPEPPTTSVSPPTTKPTTHKSPDE 741
Db 2077 tpttpttlttlttvtprpctpgtqtp--tptlttlttvtprpctpgtqtp--tpt 2132
QY 742 STELSAPPTPKALENSPKP-----GVPT--TPPATPKPEMTTAKDKTTERDL 790
Db 2133 tttvtprpctpgtqtpptlttlttvtprpctpgtqtpptlttlttvtprpctpgt 2192
QY 791 R---TTP--ETTTAAPKMT---KETATTEKTESKITAATTGTTQVTSITTTQDTTPEKITT 842

Db 2193 qtppttlttlttvtprpctpgtqtpptlttlttvtprp--tptqtppt--ltt- 2248
QY 843 KTTTLAPKVTTP--KTTTTEIEMKPEETAKPKDRATNSKATTPKPKAP--KPEPT 900
Db 2249 -ttvtprpctpgtqtpptlttlttvtprpctpgtqtpptlttlttvtprpctpgt 2307
QY 901 KKKPTMPVRKPKTTPPKMTSPMEINP--TSRIAEMLQTTTR--PNQTPNSKLVEY 956
Db 2308 qtppttlttlttvtprpctpgtqtpptlttlttvtprpctpgtqtpptlttltt 2367
QY 957 NPKSDAGAGAGETPHMLLRPHVMEVTP 986
Db 2368 ttpptpg---tqpt--tlttlttlttvt 2392

RESULT 9
AAG38942
ID AAG38942 standard; Protein: 763 AA.
XX
AC AAG38942;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 48115.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135533.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.


```

XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Misc-difference 395..396
XX FT /note="unspecified amino acids"
XX FT Misc-difference 444..446
XX FT /note="unspecified amino acids"
XX PN MO9523861-A1.
XX PD
XX PD 08-SEP-1995.
XX PE 06-MAR-1995; 95MO-CN00015.
XX PR 04-MAR-1994; 94CN-0112066.
XX PA (SHAN-) SHANGHAI BEITI BIOTECHNOLOGY CO LTD.
XX PI Gu X, Han Z, Shen Q;
XX DR MPI; 1995-320576/41.
XX DR N-PSDB; AAT04546.
XX PT New haematopoietic cell growth factor - used for treating
XX PT thrombocytopenia and hematocytopenia
XX PS Example: Page 23; 36pp; Chinese.
XX XX
XX CC This sequence represents the human megakaryocytopoietin (MPO) protein.
XX CC This sequence was purified using a carrier which can couple wheat germ
XX CC agglutinin and heparin to separate MPO. Fragments of this sequence (see
XX CC AAR80039 and AAR80040) were used to produce the amplification primers
XX CC shown in AAT04544 and AAT04545. The fragments amplified by these primers
XX CC can then be used as probes to screen human cDNA libraries for MPO cDNA.
XX CC The MPO cDNA can then be inserted into a plasmid which is used to
XX CC transform cells to produce MPO. The MPO sequence is capable of promoting
XX CC colony formation of megakaryocytes, enlarging the size of megakaryocytes
XX CC and stimulating the proliferation of multipotential stem cells. The
XX CC factor may be used for treating thrombocytopenia and hematocytopenia.
XX CC The purification method can be used to isolate MPO from human urine or
XX CC serum of patients with aplastic anaemia, and from animal blood or urine
XX CC by radiation exposing the animals to induce aplastic anaemia.
XX SQ Sequence 452 AA:

```

```

Query Match 24.3%; Score 1654; DB 16; Length 452;
Best Local Similarity 72.2%; Pred. No. 1.9e-91;
Matches 328; Conservative 3; Mismatches 21; Indels 102; Gaps 3;

```

```

QY 1 MAWKTLPYLLLLSVFVIOOVSSO----- 25
DB 1 MAWKTLPYLLLLSVFVIOOVSSO----- 25
QY 26 -----ELSCGRCFESFENGRECDCAQCKKYDKCCPDYESFCA----- 64
DB 61 krvtlaelsckgrcfesfengrecdcaqckkydkccpdyesfcaevhnpsspskcap 120
QY 65 -----EYKDNKKNNKTKKKKPPKPPVYDEAG 89
DB 121 ppsgaagtlstckrpskppnkkkkkviseeltevdkhknrtkkkpcpvpvdaag 180
QY 90 SGLDNGDFKVTTPDTSTTOHNKYSTSPKITTAKPINRPSLPNSDPSKETSIVNKKET 149
DB 181 sglngdfkvtltpdsttqhnkvstspkittakpnrpslpnsdpsketsitvnnkelt 240
QY 150 VERKETTTKQSTQSTQSTQSTQSTQSTQSTQSTQSTQSTQSTQSTQSTQSTQSTQ 209
DB 241 vekettcttkqstqstqstqstqstqstqstqstqstqstqstqstqstqstqst 300
QY 210 LTPPKPEPTTPKPEPTTPKPEPTTPKPEPTTPKPEPTTPKPEPTTPKPEPTTP 269

```

```

DB 301 ltpkpeptltpkpeptltpkpeptltpkpeptltpkpeptltpkpeptltpk 360
QY 270 PAPTPKPEPTTPKPEPTTPKPEPTTPKPEPTTPKPEPTTPKPEPTTPKPEPT 329
DB 361 paptpkpeptltpkpeptltpkpeptltpkpeptltpkpeptltpkpeptltp 409
QY 330 PREPAPTPKPEPTTPKPEPTTPKPEPTTPKPEPTTPKPEPTTPKPEPTTP 363
DB 410 prslhpkpeptltpkpeptltpkpeptltpkpeptltpkpeptltpkpept 443

RESULT 8
AAM24516
ID AAM24516 standard; Protein: 5179 AA.
AC AAM24516;
XX
XX 12-OCT-2001 (first entry)
DE C899P predicted amino acid sequence.
XX
XX Human; immunotherapy; diagnosis; colon cancer; colon tumour;
XX immunogenic; gene therapy; vaccine; colonic cancer.
XX Homo sapiens.
XX
XX WO200149716-A2.
XX
XX 12-JUL-2001.
XX
XX 29-DEC-2000; 2000MO-US35596.
XX
XX 30-DEC-1999; 99US-0476296.
XX 10-JAN-2000; 2000US-0480321.
XX 15-FEB-2000; 2000US-0504629.
XX 06-MAR-2000; 2000US-0519444.
XX 19-MAY-2000; 2000US-0575251.
XX 29-JUN-2000; 2000US-0609448.
XX 28-AUG-2000; 2000US-0649811.
XX
XX (CORI-) CORIXA CORP.
XX
XX Xu J, Lodes MJ, Secret H, Benson DR, Meagher MJ, Stolk JA;
XX King GE, Wang T, Jiang Y;
XX
XX MPI; 2001-441847/47.
XX
XX Colon tumor associated proteins and nucleic acids useful for the
XX prevention, diagnosis and treatment of colonic cancer -
XX
XX Claim 2; Page 446-462; 472pp; English.
XX
XX The present invention describes colon tumour associated proteins (I) and
XX the polynucleotides (II) that encode them. (I) have cytostatic activity.
XX (I) and (II) can be used in gene therapy and vaccine production. (I) and
XX (II) may be used in the prevention, diagnosis and treatment of diseases
XX associated with inappropriate colon tumour associated protein (TCAP)
XX expression, such as colonic cancer. For example, (I) and (II) may be
XX used to treat disorders associated with decreased expression by
XX rectifying mutations or deletions in a patient's genome that affect the
XX activity of TCAPs by expressing inactive proteins or to supplement the
XX patient's own production of them. Additionally, (II) may be used to
XX produce the TCAP proteins, by inserting the nucleic acids into a host
XX cell culturing the cell to express the protein. (II) and its
XX complementary sequences may also be used as DNA probes in diagnostic
XX polymerase chain reaction (PCR) and hybridisation assays to detect and
XX quantitate the presence of similar nucleic acids in samples, and
XX therefore which patients may be in need of restorative therapy. (I) may
XX also be used as antigens in the production of antibodies against TCAPs
XX and in assays to identify modulators of TCAP expression and activity.
XX Anti-(I) antibodies and antagonists may also be used to down regulate
XX TCAP expression and activity. The anti-(I) antibodies may also be used
XX as diagnostic agents for detecting the presence of TCAPs in samples

```


|||||
Db 721 aptcpkxpkapkelaplttckpstsdcpapltgvtgtpctpkpkeapltcpkxpkapltcpkx 780
QY 647 TAPTLKKEPAPTPPKKPAKELAPTTTGGPTSTSDKAPAPTPPKKTAATTTKREAPTPPK 706
Db 781 taptlckepapltcpkxpkapkelapltckpstsdcpapltcpkxkelapltcpkxpkapltcpkx 840
QY 707 KAPAPTPPEPPTTSEVSPPTTKEPPTTHKSPDESTPELSAEPKALENSPEKPEVPT 766
Db 841 kpaaptpceppttsevsppttckepctlnkspdestpelesepckalenspkpkypt 900
QY 767 TETPAATKPEMTTAKDKTTERDLRTPEPTTAAPKMTKEATATTEKTTESKITATTOY 826
Db 901 tktpatckpemtltakdkctterdlrttpepttaapkmktelaktteskitatttqv 960
QY 827 TSTTTODTTPPKITTLKTTTLAPKVTTTKKTTTTEIMKPEPTAKPKDRANSKATPK 886
Db 961 tstitdgttptfiktltkttlapkvttcklttetiimkpeetackpkdranskattpk 1020
QY 887 POKPTAKPKPTSTKPKKPTMPRVKPKKPTTPPKMTSTMPELNPTSRIAEAMLOTTTPN 946
Db 1021 pqkptapkkpctckpkempvrtkcttptrkmtstmpelnpstriaeamlqtttrpa 1080
QY 947 QTPNSKLVEVNPKSEDAAGBETPHMLLRPHVMEVYTPDMXYLPRVNOGIIINPMIS 1006
Db 1081 qtpnslvevnpkседagaegetphmllrphvfmpevtpdmxylprvpnglilnplms 1140
QY 1007 DETNICKGRVDDLTTLRNCTLVAFRGHYFWMLSPPSPSPARITIEWGIPSPIDTFT 1066
Db 1141 detnicgkrvddlttlrncctlvafghyfwmlspspsparritlevnglpspdtvtf 1200
QY 1067 RCNCEGTFPEFKDSQYWRFTNDIKDAGYPRIFKFGGLTGQIVAAISTAKYNNMESY 1126
Db 1201 rcncegtrffkdsqywrftndikdagypkifkgigltgqivaalstakynmpesvy 1260
QY 1127 FFRKGGSIQOYIYKOEYVOCPCGRPALANYPVGEMTOVRRRFEPRATIGSPQHTTRIQY 1186
Db 1261 ffrkggsiqoyiykqevyqcpgrpalanypvgemtqvrrrrfeteraigspqhttridy 1320
QY 1187 SPARLATYQDKGVLHNEKVSILMRGLPNVVTSAISLPNIRKPDGYDYAFASQOYINIDY 1246
Db 1321 sparltyqdkgvlhnevksilwrglpnvvtstaislpnirkpdgydyafatskdqynidv 1380
QY 1247 PSRTARATITTSQOTLSKMYNCP 1270
Db 1381 psrtaratitrsqqlskmywncp 1404
RESULT 4
AA024322
ID AA024322 standard; Protein: 1299 AA.
AC AA024322:
XX 12-OCT-2001 (first entry)
DE Human EST encoded protein. SEQ ID NO: 1847.
XX
XX Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
KW biodiversity; gene therapy; nutrition.
OS Homo sapiens.
OS
XX
XX WO200154477-A2.
XX
XX 02-AUG-2001.
XX
XX 25-JAN-2001; 2001WO-US02687.
XX
XX 25-JAN-2000; 2000US-0491404.
XX
XX 17-JUL-2000; 2000US-0617746.
PR

PR 03-AUG-2000; 2000US-0631451.
PR 15-SEP-2000; 2000US-0663870.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Werlman T;
XX
XX WPI: 2001-476164/51.
DR N-PSDB: AAH98981.
XX
XX Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use -
XX
XX Claim 20; Page 1198-1201; 1275pp; English.
PS
XX
XX The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a
CC protein of the invention.
XX
XX Sequence 1299 AA;
SQ
Query Match 90.3%; Score 6151; DB 22; Length 1299;
Best Local Similarity 89.7%; Pred. No. 0;
Matches 1161; Conservative 0; Mismatches 0; Indels 134; Gaps 2;
QY 1 MAMKTPYLLLLLSVFIQOVSQ----- 25
Db 1 mamktpllyllllsvfivqvsqdlscagrcgegyrdatcncdynqihmecpbf 60
QY 26 -----ELSCGRCFSEFRGRCDCDACCKKYDCCPDYESFCAE----- 65
Db 61 kryctaelockgrcfesfergrecddagckkydcopdyesfcaeinhnptspsskkap 120
QY 66 ----- 65
Db 121 ppsgaqtlkstktrspkppnkkkxkvieseelteebsvsengeassssssstlw 180
QY 66 -----VKDNKKKRTKKKPPKPPVYVDEAGSGGDNDFKVTTPDST 106
Db 181 kikesksnaarelqkkklykdknrtkkkpkpvpvdeagsgldngdtkvltcpdst 240
QY 107 TOHNKYSTSPKITTAPINRPSLPNSDTSKETSLVNRKETTVEKETTNNKQTSIDG 166
Db 241 tqhnkystspklttackpindrpslppnsdstsketsltnkettvetltnkqstidg 300
QY 167 KEKTSAKETOSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTTPKPEAS 226
Db 301 kektsaketosiektakdlaptskvlakptpkaetttkgpaltttpkepttpkpeas 360
QY 227 TTPKEPTTPITKSAPTTPKKEPAPTTTTSAPTTPKPEAPTTKPEAPTTTKEPAP 286
Db 361 ttpkepttpitksaptpkpkpaptttsaptpkpeaptttkpeaptttkpeaptttkp 420
QY 287 APPTTKSAPTPPKPAPPTPKKPAPTTPKPEAPTTTKEPPTTPPKPEAPTTKPEAP 346
Db 421 appttksaptpkpkpapttpkpkpapttkeappttpkpeaptttkpeaptttkp 480
QY 347 EPAPTAAPKAPPTPKPEAPPTPKPEAPTTTKEPSPPTTPKPEAPTTTKEPAPT 406
Db 481 epaptapkkpapttpkpeappttpkpeaptttkespttpkpeaptttksapttkpeapt 540
QY 407 TTKSAPTPPKESPPTTKEPAPTTPKPEAPTTPKKPAPTTPKPEAPTTTKEPAPT 466
Db 541 ttksaptpkkesppttkeappttpkpeaptttkkpaaptpkpeaptttkpeaptttkp 600
QY 467 APAPKPEPAPPTPKETAPTPKKLPTTPPEKLAFTTEKPAFTTPEELAPTTPEPTPT 526
PR

OY 1127 FFKRGSSIOOYIYKQEPYOKCPGRPALNYPYGEMTOVRRRRRPERAIGPSOTHTIRIOY 1186
 1187 SPARLAADKCVLHNEVKVSIIMRGLPNVVTSAISLPNIRKPDGKYVAFSKOQYXNIDV 1246
 Db 1261 ffrgrsg:igqylkqgvpqkcpgrpalnypvgemtyvrrrrferraiqpsqthrlrlyq 1320
 OY 1187 SPARLAADKCVLHNEVKVSIIMRGLPNVVTSAISLPNIRKPDGKYVAFSKOQYXNIDV 1246
 Db 1321 sparlalqgkylhnevkvslwrglplnvvtalslplnirkpdyqyafskodqynldv 1380
 OY 1247 PSRTARATITRSGOTLSKWNVNC 1270
 Db 1381 psrtaratitrsqgtlskwnvncp 1404
 RESULT 2
 AAB60568
 ID AAB60568 standard; Protein; 1404 AA.
 AC AAB60568;
 XX
 DT 27-APR-2001 (first entry)
 DE Human megakaryocyte stimulating factor (MSF, CACP).
 XX
 KW Human; CACP protein; campodactyly-arthropathy-coxa vara-pericarditis;
 KM MSF; megakaryocyte stimulating factor; synovial lubricant;
 KM Chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;
 KM antiarthritic.
 XX
 OS Homo sapiens.
 PN W0200107068-A1.
 XX
 PD 01-FEB-2001.
 PF 21-JUL-2000; 2000MO-US20002.
 PR 23-JUL-1999; 99US-0145328.
 PR 19-JUL-2000; 2000US-0145328.
 XX
 PA (UYCA-) UNIV CASE WESTERN RESERVE.
 PI Warman ML;
 DR WPI; 2001-182721/18.
 XX
 PT New composition comprising the campodactyly-arthropathy-coxa
 PT vara-pericarditis protein in combination with an anesthetic, useful for
 PT treating osteoarthritis, or as lubricants of tissue and joints
 PS Example 1; Page -: 34pp; English.
 XX
 CC The invention relates to a method of treating osteoarthritis via the
 CC administration of a composition comprising the campodactyly-arthropathy-
 CC coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.
 CC The composition may further comprise a local anesthetic. The composition
 CC of the invention may be administered via intra-articular or intravenous
 CC injection. The human CACP protein is identified in the invention as
 CC being megakaryocyte stimulating factor (MSF). The gene encoding
 CC CACP protein (MSF) is located on chromosome 1q25-31, and mutations in
 CC this gene are responsible for the heritable disorder campodactyly-
 CC arthropathy-coxa vara-pericarditis, in which patients have synovial
 CC hyperplasia without evidence of inflammation. CACP protein (MSF)
 CC acts as a synovium lubricant, and can be used to lubricate tissue and
 CC joints in the treatment of osteoarthritis. The composition may be
 CC applied to reduce the symptoms of osteoarthritis (e.g., joint pain,
 CC loss of range of movement or joint damage). The present sequence
 CC represents human megakaryocyte stimulating factor (MSF, CACP protein).
 CC Note: This sequence is not given in its entirety in figure 4 of the
 CC specification, although a Genbank accession number was given. This
 CC sequence was therefore obtained from Genbank (U70316).
 XX
 SQ Sequence 1404 AA;

Query Match 98.7%; Score 6727; DB 22; Length 1404;
 Best Local Similarity 90.5%; Pred. No. 0;
 Matches 1270; Conservative 0; Mismatches 0; Indels 134; Gaps 2;
 OY 1 MAMKTLPIYLILLLSVFIQGVSSQ----- 25
 Db 1 mawktlpiylilllsfvfiqgvsqdlisacagrcgegyrdctncdyncqhmecpdlf 60
 OY 26 -----ELSCRGRCFESEFRGREGCCDNOCKRYDKCCPDYSEFCAE----- 65
 Db 61 kvctaelscqgrcfeslergrecdcaqckkydkccpdyesfcaeivnlpsspskkap 120
 OY 66 ----- 65
 Db 121 psgasqtlkstktksppknkkkklkvlseeeileehsvsengessssssssstliw 180
 OY 66 -----VKDNKKNRTKKKTPPKPPVYDEAGSGLDNDFKVTTPDST 106
 Db 181 klksksaaneelqkklkvkdoknrkklkppkpvvdeagsgldngdfkvtlpldstc 240
 OY 107 TOHNKYSTSPKITTAKPINPRESLPNSDTSKETSLTVNKEETVETKETTNNKQSTDG 166
 Db 241 tqhnkystspkiltakpinpripnsdtsketsltvnkettvetkettlnkqstldg 300
 OY 167 KEKTTAKETOSIEKTSKADLAFTSKVLAKPTPKAETTTGPAALTTPKEPTTPKEPAS 226
 Db 301 kekttakektsieksakdlaptskvlakpripkaetttkgaalttpepttpkepas 360
 OY 227 TTPKEPTPTTKSAPTPPKPAATTTKSAPTTKEEAPTTTKEEAPTTTKEEAPTTTKEP 286
 Db 361 ttpkeptpttksaptppkpaatpttkesapttkeepapttkeepapttkeepapttkcp 420
 OY 287 APPTTKSAPTPKEEAPTPPKKPAATTPKEEAPTTTKEEPTTPPKKEEAPTTKEEAPTTK 346
 Db 421 apttksaptpkeepaptppkkpaattpkeepaptttkesapttkeepapttkeepapttk 480
 OY 347 EPAPTPAKKPAATTPKEEAPTTTKEEAPTTTKEEAPTTTKEEAPTTTKEEAPTTTKEP 406
 Db 481 epaptpakpapttkkeepaptttkesapttkeepaptttkesaptttkesapttkcpap 540
 OY 407 TTKSAPTPKEEPTPTTKEEAPTTTKEEAPTTTKEEAPTTTKEEAPTTTKEEAPTTTKEP 466
 Db 541 ttksaptpkeepspttkkeepaptttkesaptttkesaptttkesaptttkesapttkcp 600
 OY 467 APAPKEEAPPTPKETAPTTPKKLTPTTPEKLAFTTPEKAPTTPEELATTPEEPTPT 526
 Db 601 aptapkeepaptpkelaftpttkkltpttpekelapttpeelaptppeelaptppeept 660
 OY 527 PEEAPPTPKAAANTPKKEEAPTTTKEEAPTTTKEEAPTTTKEEAPTTTKEEAPTTTKEP 586
 Db 661 peepaptpkkaantpkkeepaptttkesaptttkesaptttkesaptttkesapttkcp 720
 OY 587 APPTPKKPAEKLAPTTTKEEPTTTSKPAATTPPKGTAATTPPKAPTTTKEEAPTTPKG 646
 Db 721 aptpkpkpaelaptpttkeepssttskpaattpkgtattppkaptttkesapttkcp 780
 OY 647 TAPPTLKPAATTPPKKPAKLAFTTTPKSTTSKPAATTPPKETAATTPPKKEEAPTTPK 706
 Db 781 tapptlkpaattpkkpaelaptttkgsttskpaattpketaattppkkeepapttkcp 840
 OY 707 KPAPTPPEPTTSEVSTPTTKEPTTIHKSPEDESTPELSAEPTPKALNSKPEGPVT 766
 Db 841 kpaftppepttsevspttkepttihkspestpeelsaeptpkalnskpepvyt 900
 OY 767 TKTPATKPEMTTAKDKTTERDLRTPTETTYAAPKMTKEPATTTETKESKITATTTQV 826
 Db 901 tktpatkepmttakdkterdlrtptettaapkmketaattekteskitaattqv 960
 OY 827 TSTTTQDTTPPKKITLTKTTTTLAPKVTTTKKTTTTEIINNKKEETAKPKDRATNSKATTPK 886
 Db 961 tsttqdttpkkitlktttlapkvtttkktttteimnkpeetakpkdratnskattpk 1020

```
FT      Region      1266..1331
FT      /label= Exon_X
FT      Region      1331..1373
FT      /label= Exon_XI
FT      Region      1373..1404
FT      /label= Exon_XII
XX      W09213075-A.
XX      06-AUG-1992.
XX
XX      17-JAN-1992. 92WO-US00433.
XX
XX      18-JAN-1991. 91US-0643502.
XX      10-SEP-1991. 91US-0757022.
XX      (GEMV ) GENETICS INST INC.
XX
XX      Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K;
XX      WPI; 1992-284660/34.
XX      N-PSDB; AAQ27223.
XX
XX      New human mega-karyocyte stimulating factors - for treating
XX      immune deficiencies, cancer, exposure to radiation or drugs,
XX      bacterial and viral infections, etc.
XX
XX      Claim 1, 2 and 3; Fig 1; 87pp; English.
XX
XX      The sequence given is a full length translation from the megakaryocyte
XX      stimulating factor (MSF) precursor. The sequence covered by exons II,
XX      III and IV encodes megakaryocyte stimulating factor (MSF). This
XX      sequence is modified by the addition of an N-terminal sequence encoding
XX      a secretory leader, an initiating methionine preceding exon II and a
XX      terminating codon following exon IV. The cDNA sequence given contains
XX      sequences derived from human megakaryocyte colony stimulating factor
XX      (meg-CSF). Exon I contains the initiating methionine, and encodes a
XX      classical mammalian protein secretion signal sequence. The sequence
XX      terminating the original meg-CSF includes exons II-IV and is thought to
XX      terminate in the region between amino acid residues 134 - 147. The
XX      yield a family of mRNAs each encoding a different MSF protein. Exons
XX      V and VI are thought to be related to the activity of the factor and
XX      are also implicated in the stability, folding and processing of the
XX      molecule. These exons are also thought to play a role in the observed
XX      synergy of MSF with other cytokines. Exons V - XII are believed to be
XX      implicated in the processing or folding of the appropriate structure of
XX      the resulting factor, ie. one or more of these exons may contain
XX      sequences which direct proteolytic cleavage, adhesion, organisation of
XX      the cellular matrix or extracellular matrix processing. Both naturally
XX      occurring and non-naturally occurring MSF's may be characterised by
XX      various combinations of alternatively spliced exons from this sequence,
XX      members of the MSF family.
XX
XX      Sequence 1404 AA:
XX
Query Match      98.7%; Score 6727; DB 13; Length 1404;
Best Local Similarity 90.5%; Pred. No. 0;
Matches 1270; Conservative 0; Mismatches 0; Indels 134; Gaps 2;
QY      1 MANKTPIYLLLSLVFVIOQVSSQ-----
DB      1 mawkltpllyllllsvfvlqvsagqlssacagrcgegyrdatcndyncqhyemccpdf 60
QY      26 -----ELSCGRCFESFERGRCDCDAQCRRYDKCCPYESFCAE----- 65
DB      61 kvtcaelackgrcfesfergrecddagckkydkccpdyesfcaeivhmpstpskkap 120
QY      66 ----- 65
DB      121 ppsgaagtlkstktspkpnkktkkyvlseeeiteehsvsengeassssssssstliw 180
```

```
QY      66 -----VKDNKKNRTRKKRPPEKPPVVDAGSGLDNGDEVTTPPTST 106
DB      181 ktkssksaanreqlkklkvdknknrckkrcpckppvvdgaagslndgdcvkvtptstc 240
QY      107 TOHNRVSTSPKITTAKPPIPNRPSLPPNSDTSKESLTVNKKETTYETKETTNNKQSTG 166
DB      241 tqhkvstspklttckapnprpslppnsdtskeslvtvnkeltvckettlnkqstcdg 300
QY      167 KEKTTSAKETOSIEKTSKADLAPTSKVLAKEPPKAETTTGKGPALTTPKEPPTPKEPAS 226
DB      301 kektsaketqsietsakdlaptskvlakpkaetttgkpaltpkepttckepas 360
QY      227 TTPKEPPTTTKSAPTTPKEPAPTTTTSAPPTPKKEPAPTTTKEPAPTTTKEP 286
DB      361 ttpkepptttksapttpkpapttttsapptpkkepaptttkepaptttkep 420
QY      287 APTTTSAPPTPKKPAPPTPKKPAPTTTPKEPAPTTPKKEPAPTTPKKEPAPTT 346
DB      421 apttksaptpkcpaptpckkpaptpckepaptpckepaptpckepaptpck 480
QY      347 EPAPTAKKPPAPTTPKKPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 406
DB      481 epaptakkpaptpckpapttkpapttkpapttkpapttkpapttkpapttk 540
QY      407 TTKSAPTTPEKSPPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 466
DB      541 ttksaptpkcpaptpckpapttkpapttkpapttkpapttkpapttkpapttk 600
QY      467 APTAPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 526
DB      601 aptapkepapttkcpaptpckpapttkpapttkpapttkpapttkpapttk 660
QY      527 PEEPAPTPKAAAPNTPKKPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 586
DB      661 peepaptpkkaapntpkkpaptptkpapttkpapttkpapttkpapttkp 720
QY      587 APTTPKPKPKELAPTTTKEPSTSDKRAPTTTKEPAPTTTKEPAPTTTKEPAP 646
DB      721 apttpkpkpkelaptpckpapttkpapttkpapttkpapttkpapttkp 780
QY      647 TAPTTKEPAPTTPKKPAPKEAPTTTNGPSTGSDKRAPTTTKEPAPTTTKEPAP 706
DB      781 tapttkepaptpkpkpapkapttkpapttkpapttkpapttkpapttkp 840
QY      707 KPAPTPEPPTTSEVSTPTTTPKPTTIHKSPEDESTPELSAEPKALENSPKPEPVPT 766
DB      841 kpaptpeppttsevspttpttkppttihspsdestpelsaepkalenspek 900
QY      767 TTPPATRPEMTTAKDKTTERDLRTTPEPTTAAPKMKETATTTKTTESKITATTT 826
DB      901 ttppatrpemttakdkttterdlrttpeptttaaapkmetatttktteskita 960
QY      827 TSTTQODTTPPKITITLTKTTTTLAPKVYTTKKTIITTEIMNKPEETAKPRDRATNSKAT 886
DB      961 tsttqodttpkkititlkttttlapkvtyttkktiitteimnkpeetakprdr 1020
QY      887 POKPTAKKRPSTGKPKTPMPPRVKRPKTPPRKMTSTMPLENTSTSLAAMLOTTTRPN 946
DB      1021 pkpptakkrpstgkpktpmpprvkrpktpprkmtstmplentpstsilaamlot 1080
QY      947 QTPNSKLVEVNPKSEDAAGAGETPPHMLLRPHVMPREVTDMVDLPRVPMQGIINPMLS 1006
DB      1081 qtpnslkvevnpskedagaggetpphmlrphvmprevtdmvdprvpngqiln 1140
QY      1007 DETNIGKGRPVGDGTTJTRNGTLVAFRGHYPMMLSPSPSPARITRWGIVSPIDTVFT 1066
DB      1141 detnigkgrpvgdgttjtrngtlvafrghyfmlspspsparitrwgivspidt 1200
QY      1067 RCNCEGKTFPFFKDSQWRFTNDIKDAGYPRPIFKGSGGLGQIVAAALSTAKYKNMPESTY 1126
DB      1201 rcncegkttffkdsqywrftndikdagypkpifkgsyglgtgqlvaalstak 1260
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:24:31 ; Search time 138.84 Seconds
(without alignments)
677.565 Million cell updates/sec

Title: AA3
6814
Sequence: 1 MAWKLPYILLLSVFVQ.....AAATIRSGQTSKRWYNCP 1270

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SID58/gcgdata/geneseq/AA1980.DAT.*
2: /SID58/gcgdata/geneseq/AA1981.DAT.*
3: /SID58/gcgdata/geneseq/AA1982.DAT.*
4: /SID58/gcgdata/geneseq/AA1983.DAT.*
5: /SID58/gcgdata/geneseq/AA1984.DAT.*
6: /SID58/gcgdata/geneseq/AA1985.DAT.*
7: /SID58/gcgdata/geneseq/AA1986.DAT.*
8: /SID58/gcgdata/geneseq/AA1987.DAT.*
9: /SID58/gcgdata/geneseq/AA1988.DAT.*
10: /SID58/gcgdata/geneseq/AA1989.DAT.*
11: /SID58/gcgdata/geneseq/AA1990.DAT.*
12: /SID58/gcgdata/geneseq/AA1991.DAT.*
13: /SID58/gcgdata/geneseq/AA1992.DAT.*
14: /SID58/gcgdata/geneseq/AA1993.DAT.*
15: /SID58/gcgdata/geneseq/AA1995.DAT.*
16: /SID58/gcgdata/geneseq/AA1996.DAT.*
17: /SID58/gcgdata/geneseq/AA1997.DAT.*
18: /SID58/gcgdata/geneseq/AA1998.DAT.*
19: /SID58/gcgdata/geneseq/AA1999.DAT.*
20: /SID58/gcgdata/geneseq/AA2000.DAT.*
21: /SID58/gcgdata/geneseq/AA2001.DAT.*
22: /SID58/gcgdata/geneseq/AA2001.DAT.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6727	98.7	1404	AA26049	MSF precursor. Sy
2	6727	98.7	1404	AA26056	Human megakaryocyt
3	6727	98.7	1404	AA26056	Human megakaryocyt
4	6151	90.3	1299	AA26322	Human EST encoded
5	3484	51.1	902	AA26378	Human MSF ortholo
6	1707.5	25.1	472	AA26056	Bovine MSF ortholo
7	1654	24.3	452	AA26056	Human megakaryocyt
8	1171	17.2	5179	AA26056	C899P predicted am
9	981	14.4	763	AA26056	Arabidopsis thalila
10	950	13.9	1664	AA26056	C. thermocellum OI
11	776.5	11.4	1325	AA26056	Peptide #2327 enco

12	715	10.5	763	18	AA261852
13	627	9.2	4412	21	AA263666
14	612	9.0	572	18	AA261855
15	547.5	8.0	844	7	AA260570
16	542	8.0	807	21	AA264467
17	520.5	7.6	788	21	AA264466
18	508	7.5	1837	21	AA261126
19	506.5	7.4	744	9	AA262975
20	496.5	7.3	2971	21	AA2641231
21	489.5	7.2	3118	22	AA2650363
22	489.5	7.2	3118	22	AA2650362
23	489	7.2	1721	21	AA2611727
24	488.5	7.2	826	13	AA266042
25	488	7.2	617	22	AA2616458
26	488	7.2	957	21	AA264513
27	488	7.2	957	22	AA264513
28	488	7.2	1721	21	AA2648299
29	485	7.1	1127	22	AA265541
30	476.5	7.0	652	9	AA262974
31	467.5	6.9	511	22	AA264883
32	467.5	6.9	511	22	AA2627312
33	467.5	6.9	511	22	AA262607
34	467.5	6.9	750	20	AA265477
35	455.5	6.7	378	12	AA264162
36	450.5	6.6	2870	21	AA265559
37	446.5	6.5	3178	21	AA265556
38	443	6.5	751	16	AA260839
39	443	6.5	2819	22	AA2635408
40	442.5	6.4	1610	21	AA262870
41	437.5	6.3	1610	21	AA269648
42	429	6.3	2665	22	AA261533
43	428.5	6.3	2665	22	AA261533
44	428.5	6.3	2665	22	AA261533
45	428.5	6.3	2665	22	AA26950

ALIGNMENTS

RESULT 1	AA26049	standard: Protein; 1404 AA.
ID	AA26049	
AC	AA26049	
XX	02-FEB-1993 (first entry)	
DT	MSF precursor.	
XX		
DE		
XX	Megakaryocyte colony stimulating factor; secretion signal; meg-CSF;	
KW	stability; proteolytic cleavage; adhesion; alternative splicing.	
KW		
XX	Synthetic.	
OS		
XX		
PH	Key	Location/Qualifiers
FT	Region	1..26 /label= Exon_I
FT	Region	26..67 /label= Exon_II
FT	Region	67..107 /label= Exon_III
FT	Region	107..157 /label= Exon_IV
FT	Region	157..200 /label= Exon_V
FT	Region	200..1141 /label= Exon_VI
FT	Region	1411..1166 /label= Exon_VII
FT	Region	1166..1212 /label= Exon_VIII
FT	Region	1213..1266 /label= Exon_IX
FT	Region	

Mycobacterium tube
Sequence g1/101742
Mycobacterium tube
Sequence of the Fa
Amino acid sequenc
Amino acid sequenc
Cryptosporidium pa
Bioadhesive precu
Human ORFX ORF95
Human SRCAP. Homo
Human SRCAP. Homo
Portion of Cryptos
P. yoelii SSP ant
Peptide #2892 enco
Peptide #2892 enco
Human MUC11 polype
C900P predicted am
Cryptosporidium pa
Human protein sequ
Bioadhesive precu
Peptide #1317 enco
Peptide #1349 enco
Peptide #1289 enco
C. albicans Rb1 p
PRP 378. Triticum
Caenorhabditis ele
Caenorhabditis ele
Japanese sea musse
Human 07CG27 gene
EYFP-DEVD-MAP4-EBF
Bifunctional caspa
Peptide #967 enco
Peptide #967 enco

Mon Apr 29 08:35:31 2002

us-09-556-246-1_copy_200_1140.rag

b. 2 - 3

Page 16

XX	Sequence of the Falciparum Interspersed Repeat Antigen
DE	(FIRA).
XX	
KW	Malaria vaccine; antigen; epitope.
OS	
XX	Plasmodium falciparum.
PN	
XX	WO601802-A.
PD	
XX	27-MAR-1986.
PX	
PF	11-SEP-1985;
XX	85MO-0006960.
PR	11-SEP-1984;
XX	84AU-0007067.
PR	11-SEP-1984;
XX	84AU-0007066.
PR	10-SEP-1985;
XX	85AU-0047326.
PA	
XX	(HALL-) HALL INST MED RES.
PI	
DR	Kemp DJ, Anders R, Coppel RL, Brown G, Saint RB, Cowman AF;
XX	
DR	WPI: 1986-094065/14.
XX	N-PSDB: AAN60473.
PT	
XX	DNA coding for Plasmodium falciparum antigens - expressing
PT	poly:peptide(s) having antigenicity of RESA or FIRA antigens of P
XX	falciparum
PS	
XX	Disclosure: Fig 7; 55pp; English.
CC	
CC	The inventors claim a novel DNA molecule which comprises a
CC	nucleotide sequence corresp. to all or a portion of the base
CC	sequence coding RESA (AAN60472) or FIRA (AAN60473). RESA and FIRA have
CC	antigenicity suitable for providing protective immunity against
XX	Plasmodium falciparum malarial infections.
XO	
Sequence	844 AA;

[illegible][illegible]

RESULT	15	
AA54467		
ID	AA54467	standard; Protein; 807 AA.
XX		
AC	AA54467:	
XX		
DT	25-APR-2000	(first entry)
DE		
XX	Amini acid sequence of intestinal	Insect mucin isoform IIM22.
XX	Intestinal	Insect mucin; IIM; isoform; IIM14; peritrophic membrane
KW	Insect pest; transgenic plant; IIM22;	bio-engineered pesticide;
KW	enhancin; Insect vector.	
XX		
OS	Trichoplusia ni.	
XX		
FT	Key	Location/Qualifiers
FT	Peptide	1..25
FT		/note= "signal peptide"
FT	Misc-difference	36
FT		/note= "unspecified amino acid encoded by TGC"
FT	Misc-difference	42
FT		/note= "unspecified amino acid encoded by CAC"
FT	Misc-difference	43
FT		/note= "unspecified amino acid encoded by TGS"
FT	Misc-difference	47
FT		/note= "unspecified amino acid encoded by CAC"
FT	Misc-difference	685
FT		/note= "His encoded by CCT"
PN	WO967373-A2.	
PD	29-DEC-1999.	
PF	23-JUN-1999;	99WO-US14220.
XX	24-JUN-1998;	98US-0103429.

QY 132 PTPKAEETTTGSPALTTPEKPTTTPEKPASTTPEKPTTTIKASPTTPEKPAETTTKSA 191
 Db 3 PVP-----APRALPRLPAPPAEPKAPKPFPPAP-----PAPCMMLVSAAP 46
 QY 192 TTPKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKE 251
 Db 47 PCP---PAP---PAPKPKSKAPFPVPVPPAPPAELAPLP---PAP-----PAPRE 90.
 QY 252 PAPTTPKEPTPTT-----PKEPAPTTPKEPAPTTPK---EPAPAPK---KPAPTTPKEPAPTTPK 305
 Db 91 SRAIPPCPPPPVVLIDPPPEAAPPPAPNSPFPFPKPKFPVPPVP---PVPNSP 148
 QY 306 EP---APTTPKEPPTTPKEPAPTTPKSAPTTPKKEPAPTTPKSAPTTPKEPPTTPKKEPAP 363
 Db 149 FPFIPPAALPPAP-----PAPPLANSPLPPAPLPAGT---PPAPWPVPPAPKSKPA 201
 QY 364 TTPKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPK 419
 Db 202 SPPTPPAP-----PMPALPMEFPPLPVPVPPDLSKECPAPPPAPPLPVPVPLP 256
 QY 420 TTPKELTPPTPEKLP-----TTPKEPAPTTPPEELAPPTPEEPTTPPEEPAPTTP-- 469
 Db 257 PVPKLIAPPAP---APPVAVAAVIVAPCPPLPPLPMPHPAPPAAPVPPVPLAPLPNSHP 313
 QY 470 ---PKAAPPTPEKAP-----PAPPAAPVPPVPLAPLPNSHP 313
 Db 314 PAPPAPVPPVPLAPLPISGRVSVKGSFCLISFCRCVSGVLAGALNSPISPSPL 373
 QY 484 ---PTPKPAPTTPKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPK 533
 Db 374 TTTTAPLAPLAPLPPPLPPLPPLPPLPPLPPLPPLPPLPPLPPLPPLPPLPPLPPLPPL 431
 QY 534 LAPPTTKEPTSTSDKPAPTTPKGAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPK 587
 Db 432 ---PAPKPKVPLPPLAPLAPPEPK---LVPVLPPGSCCPSEKPNPAPPEPPEKSSPAPLP 488
 QY 588 KEPAFTTP---KKPAKELAPTTTKGP---TSTSDKPAPTTPKKEPAPTTPKKEPAPTTP 640
 Db 489 APPAPMPSAVIVPPAPPLPAPPAPPAAPPAAPLAPPAPSPALPLCPPLPSPAPNSP 548
 QY 641 KKPAPTTPPTPTTSEVSTPTTKEPTTIHKSP---DSTPELSAEPKPA---LENSP 694
 Db 549 ---PAPAPPPPLPPLPPLPPLPPLPPLPPLPPLPPLPPLPPLPPLPPLPPLPPL 604
 QY 695 KEGVPTTPKPAATKPEMTTAAKDKTTERDLKTPPETTAAKPKMTKETATTEKTESKI 754
 Db 605 ---PAPAPPPAPPAALPVPNPA-----PPLPAAPK-----SPAL 637
 QY 755 TATTTQVSTTTQDTTPFKITTLKTTTLAKVTTTKTTTTEIMKPEETAKPKDRATN 814
 Db 638 PAAPPAPPPAPPVATCP-----PPAPPPAPPA 665
 QY 815 SKATTPKPOK---TKAPKPTSTKPKKPMVRRKPTTPPRKMTSTV-----PELNP 865
 Db 666 SMLAPPAPPPPLPPLPPLPPLPPLPPLPPLPPLPPLPPLPPLPPLPPLPPLPPL 725
 QY 866 TSRIAEMLQTTTRPMOTPNKSLVEVNPXSEDAAGAGET 905
 Db 726 VPQAPLAPLPINGRVPFATKNSLI-----GSSSGAT 756

XX OS Unidentified.
 XX PN W09960164-A1.
 XX PD 25-NOV-1999.
 XX PE 14-MAY-1999; 99NO-US11066.
 XX PR 15-MAY-1998; 98US-0085673.
 XX PA (QUAR-) QUARK BIOTECH INC.
 XX PI Elinat P, Mor O, Skalter R, Feinstein E, Faerman A;
 XX DR WPI; 2000-053304/04.
 XX PT Identification of stress induced genes for determining risk and
 XX PT preventing, treating or controlling osteoporosis
 XX PS Claim 32; Fig 6A-R; 308pp; English.
 CC The present sequence is obtained from a clustal X alignment with
 CC protein 608. Protein 608 was identified using the method of the invention
 CC after subjecting rat osteoblasts to mechanical stress. Expression of the
 CC 608 gene was found to be upregulated by about 3-fold in cells subjected
 CC to mechanical strain. The specification describes a method for the
 CC identification of genes responsive to a specific mechanical stress. The
 CC method comprising applying the mechanical stress to an organism (tissue
 CC or cells comprising bone cells), isolating the specific cellular
 CC mRNA in comparison with control samples. The method is used to identify
 CC genes whose expression is responsive to a specific stress. The identified
 CC genes are employed in determining risk associated with a physiological or
 CC disease state. The risk determination methods are used for testing a
 CC medication for gene therapy. These medications, or genes identified by
 CC the method of the invention, are used for treating, preventing or
 CC controlling a physiological or disease state (especially osteoporosis or
 CC bone density or other factors causing or contributing to osteoporosis or
 CC its symptoms or other conditions involved in mechanical stress or its
 CC lack. The methods can also be used for advancing research or studies in
 CC bone development.
 CC XX
 SQ Sequence 4412 AA;
 Query Match 12.5%; Score 625.5; DB 21; Length 4412;
 Best Local Similarity 26.2%; Pred. No. 5.2e-27;
 Matches 225; Conservative 86; Mismatches 379; Indels 169; Gaps 40;
 QY 99 TDGKEKTTSAKETOSIEKT-----SAKDIAPTSVLAKPPPKAETTTGSPALT----- 146
 Db 3515 tdtirgilqrveehvrekhrvlefeaevefekpkapp-----kgsisekllp 3566
 QY 147 TPKETPTTP---KEPASTTPKEPTTPPTTTPKKEP---APTTPKSAPTTPKKEPAPTTPKE 204
 Db 3567 pkkpcklvprkepkavpevkkllveekvrvpeesrvpcklvpevlpkvevp----- 3621
 QY 205 PAPTTPKEPAPTTPKKEPAPTTPKSAPTTPKKEPAPTTPK---PAPTTPKEPAPTTPKKEP 260
 Db 3622 ---ekkvvpakpkpeap-----pkkvpeapkevevpevppkpkvepckkvpev 3672
 QY 261 TPTTPKEPAPTTPKKEPAPTTP-----KEPAPAPKPPAPPTTPKKEPAPTTPKKEPAPTTP 313
 Db 3673 kaavpekavpealpkpeespppevfepespsapkp---kvevppvrvpevkevepkv 3731
 QY 314 EDSPTTPKKEPAPTTPKSAPTTPKKEPAPTTPKSAPTTPKKEPAPTTPKKEPAPTTP 373
 Db 3732 paap---pkkpevtvkv---vpeapkevevpevpp---pkkpevppck-----vpevkvav 3782
 QY 374 PKKPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPK 429
 Db 3783 pekvp-----ealpkpkeespppevfepespsapkp---kvevppvrvpevkevepkv 3838

Db 2360 pltttttvtptptgt---tqtp--ttptttttvtcp 2392

RESULT 8
AAG38942
ID AAG38942 standard; Protein: 763 AA.
XX
AC AAG38942;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 48115.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PE 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134286.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140635.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.

AA24516
ID AA24516 standard; Protein; 5179 AA.
XX
AC AA24516;
XX DT 12-OCT-2001 (first entry)
XX DE C899P predicted amino acid sequence.
XX
KW Human; immunotherapy; diagnosis; colon cancer; colon tumour;
KW immunogenic; gene therapy; vaccine; colonic cancer.
XX
OS Homo sapiens.
XX WO200149716-A2.
XX PN 12-JUL-2001.
XX PD
XX PF 29-DEC-2000; 2000WO-US35596.
XX PR 30-DEC-1999; 99US-0476296.
XX PR 10-JAN-2000; 2000US-0480321.
XX PR 15-FEB-2000; 2000US-0504629.
XX PR 06-MAR-2000; 2000US-0519444.
XX PR 19-MAY-2000; 2000US-0575251.
XX PR 29-JUN-2000; 2000US-0609448.
XX PR 28-AUG-2000; 2000US-0649811.
XX PA (CORI-) CORIXA CORP.
XX PI Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
PI King GE, Wang T, Jiang Y;
XX WPI: 2001-441847/47.
XX DR
XX PT Colon tumor associated proteins and nucleic acids useful for the
PT prevention, diagnosis and treatment of colonic cancer -
XX
XX Claim 2; Page 446-462; 472pp; English.
XX
XX The present invention describes colon tumour associated proteins (I) and
XX the polynucleotides (II) that encode them. (I) have cytosolic activity.
XX (I) and (II) can be used in gene therapy and vaccine production. (I) and
XX (II) may be used in the prevention, diagnosis and treatment of diseases
XX associated with inappropriate colon tumour associated protein (TCAP)
XX expression, such as colonic cancer. For example, (I) and (II) may be
XX used to treat disorders associated with decreased expression by
XX rectifying mutations or deletions in a patient's genome that affect the
XX activity of TCAPs by expressing inactive proteins or to supplement the
XX patients own production of them. Additionally, (II) may be used to
XX produce the TCAP proteins, by inserting the nucleic acids into a host
XX cell culturing the cell to express the protein. (II) and its
XX complementary sequences may also be used as DNA probes in diagnostic
XX polymerase chain reaction (PCR) and hybridisation assays to detect and
XX quantitate the presence of similar nucleic acids in samples, and
XX therefore which patients may be in need of restorative therapy. (I) may
XX also be used as antigens in the production of antibodies against TCAPs
XX and in assays to identify modulators of TCAP expression and activity.
XX Anti-(I) antibodies and antagonists may also be used to down regulate
XX TCAP expression and activity. The anti-(I) antibodies may also be used
XX as diagnostic agents for detecting the presence of TCAPs in samples
XX (e.g. by enzyme linked immunosorbent assay (ELISA)). AA1284460 to AA129512
XX and AA24494 to AA24523 represent nucleotide and amino acid sequences
XX given in the exemplification of the present invention.
XX
XX Sequence 5179 AA;

QY	21	DEAGSGLDNGDFK-----VTPDSTTQH-NKVSTSPK-----	52
----	----	---	----

[illegible][illegible]

CC human MSF.
 XX
 SQ Sequence 1404 AA;

Query Match 100.0%; Score 5011; DB 22; Length 1404;
 Best Local Similarity 100.0%; Pred. No. 1.6e-273;
 Matches 941; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VKDNKKNRRTKKKKPPKPPVVDGAGSGLDNGDFKVTPTPTSTGTHNKKVSTSPKITTAKPIN 60
 |||||||
 DB 200 vkdnknrtkkkppkppvvdagsgldngdfkvtptdstctqhnkvsstkpkittakpin 259
 |||||||
 QY 61 PRPSLPNSDTSKETSILVVKETKETTNNKOTSTGKEKTSAKETOSIEKTSK 120
 |||||||
 DB 260 prpslpnsdtsketsilvkvkettettnkqstlgketsktsaketqsieksak 319
 |||||||
 QY 121 DLAPTSKVLAKPPKAPATTGSPALTPPKETPTTPKEPASTTPKEPTPTTIKSAPTTPK 180
 |||||||
 DB 320 dlaptskvialkppkaetctckgpalcttpktpktpkpssttpktpktpktsaplttk 379
 |||||||
 QY 181 EPAPTTKTSAPTTPEKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAPATT 240
 |||||||
 DB 380 epapttksapttpekapattpekapattpekapattpekapattksaplttkapaplt 439
 |||||||
 QY 241 PKKAPTTPEKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAP 300
 |||||||
 DB 440 pkkapttpekapattpekapattpekapattpekapattpekapattkpkapltkpe 499
 |||||||
 QY 301 PTPKEPAPTTPEKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAPATTPEK 360
 |||||||
 DB 500 ptpkepattpekapattpekapattpekapattpekapattksaplttkapaplttk 559
 |||||||
 QY 361 PAPPTPEKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAPATT 420
 |||||||
 DB 560 papltpekapattpekapattpekapattpekapattpekapattkpkapaplttkcap 619
 |||||||
 QY 421 TPKKLTPTPEKLAPTPEKAPATTPEELAPTTPEEPPTPEEPAPTTPEAPAAAPNTPE 480
 |||||||
 DB 620 tpkkltptpeklaptpckepapltpeelaptpceepptceepapltkpaapnptkpe 679
 |||||||
 QY 481 PAPPTPEKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAPATT 540
 |||||||
 DB 680 papltpekapattpekapattpekapattpekapattpekapattkpkapaplttk 739
 |||||||
 QY 541 EPSTSTSKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAPATT 600
 |||||||
 DB 740 epsttsdkapattpekapattpekapattpekapattpekapattclkepapltcpk 799
 |||||||
 QY 601 KELAPTTKSGPTSTSDKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAPATT 660
 |||||||
 DB 800 kelpatltksgptstsdkapattpekapattpekapattpekapattpeprrptsevt 859
 |||||||
 QY 661 PTTTKEPTTIKSPDESPELSAETPKALNSKPEGVPTTKPAAKKPEMTTAKAKXT 720
 |||||||
 DB 860 ptttkepttlkspdespelsaetpkalnskpegvpttkpaaekpentltakakt 919
 |||||||
 QY 721 TERDLRTPEPTTTAPKATKETAATTEKTESKITAATTQVYSTTQDTPTFFKITTLKTT 780
 |||||||
 DB 920 terdlrtpeptttapkatketaattekteskitattqyvssttqdtptffkitlcltt 979
 |||||||
 QY 781 TLAPKVTITTKTITTTETIMNKRDEETAKPKDRATNSKATTPKQKPTKAPKPKTSKKPKT 840
 |||||||
 DB 980 tlapkvtttkktttettelmnkpeetapkdratnaskattpkqpktpkpkpkscltkpct 1039
 |||||||
 QY 841 MPRVAKKPTTTPPKAMTSMPELNTSRIAAEMLOTTTRPNQTPMSKIVENVKPSKEDAGG 900
 |||||||
 DB 1040 mprvakpcttppkamtstmpelntsrilaemloTTTRPNQTPMSKIVENVKPSKEDAGG 1099
 |||||||
 QY 901 AEGETPHMLRPHVMPPEVTPDMOYLPRVNOGIIINPMLS 941
 |||||||
 DB 1100 aegetchmlrphvmppevtpdmoylprvnggiiinpmlls 1140

RESULT 5
 AAB29778
 ID AAB29778 standard; protein; 902 AA.
 XX
 AC AAB29778;
 XX
 DF 28-FEB-2001 (first entry)
 XX
 DE Human MSF-derived tribonecclin.
 XX
 XX Human tribonecclin; MSF; megakaryocyte stimulating factor;
 KW alternative splicing; joint boundary lubricant; O-linked oligosaccharide;
 KW osteoarthritis; tribosupplementation; tissue adhesion inhibition;
 KW friction coefficient reduction; gene therapy; antiarthritic;
 KW osteopathic.
 XX
 OS Homo sapiens.
 XX
 PN W0200064930-A2.
 XX
 PD 02-NOV-2000.
 XX
 PF 24-APR-2000; 200OMC-US10953.
 XX
 PR 23-APR-1999; 99US-0298970.
 XX
 PA (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.
 XX
 PI Jay GD;
 XX
 DR WPI; 2001-024673/03.
 XX
 PT Novel tribonecclin polypeptide useful as lubricant for treating
 XX osteoarthritis, comprises O-linked lubricating moiety -
 PS Disclosure; Fig 1; 47pp; English.

CC The invention relates to a human tribonecclin which is a product of
 CC alternative splicing of the human MSF (megakaryocyte stimulating factor)
 CC gene. The tribonecclin has at least one O-linked oligosaccharide
 CC lubricating moiety and has a polypeptide sequence comprising 1-76
 CC repeats of a motif having at least 50% identity to the sequence KEAPPT
 CC (AAB29774). The invention also relates to a nucleic acid encoding a
 CC human MSF-derived tribonecclin; a biocompatible composition comprising a
 CC human tribonecclin for inhibiting tissue adhesion formation; and a method
 CC of diagnosing osteoarthritis or a predisposition to osteoarthritis by
 CC measuring the amount of MSF or its fragment in a biological sample of a
 CC mammal, wherein an increased amount of MSF compared to a control
 CC indicates the presence of or predisposition to developing
 CC osteoarthritis. The tribonecclin and DNA encoding it are useful in the
 CC treatment of osteoarthritis, where they may be used for lubricating
 CC mammalian joints, such as articulating joints of humans, dogs or horses.
 CC The tribonecclin, when formulated as a membrane, foam, gel or fibre, is
 CC useful for inhibiting adhesion between two surfaces such as the injured
 CC tissues of a mammal, where the injury is caused by a surgical insertion
 CC or trauma, or an artificial device e.g., an orthopaedic implant. In
 CC particular, one of the surfaces is pericardial tissue. DNA encoding a
 CC tribonecclin may be used in gene therapy. The present sequence represents
 CC a substantial portion of a human MSF-derived tribonecclin.

XX
 SO Sequence 902 AA;

Query Match 56.1%; Score 2812.5; DB 22; Length 902;
 Best Local Similarity 71.4%; Pred. No. 2.3e-150;
 Matches 672; Conservative 28; Mismatches 80; Indels 161; Gaps 65;

QY 1 VKDNKKNRRTKKKKPPKPPVVDGAGSGLDNGDFKVTPTPTSTGTHNKKVSTSPKITTAKPIN 60
 |||||||
 DB 1 vkdnknrtkkkppkppvvdagsgldngdfkvtptdstctqhnkvsstkpkittakpin 60
 |||||||
 QY 61 PRPSLPNSDTSKETSILVVKETKETTNNKOTSTGKEKTSAKETOSIEKTSK 120

CC Injection. The human CACP protein is identified in the invention as
 CC being megakaryocyte stimulating factor (MSF). The gene encoding
 CC CACP protein (MSF) is located on chromosome 1q25-31, and mutations in
 CC this gene are responsible for the heritable disorder camptodactyly-
 CC arthropathy-coxa vara-pericarditis, in which patients have synovial
 CC hyperplasia without evidence of inflammation. CACP protein (MSF)
 CC acts as a synovium lubricant, and can be used to lubricate tissue and
 CC joints in the treatment of osteoarthritis. The composition may be
 CC applied to reduce the symptoms of osteoarthritis (e.g., joint pain,
 CC loss of range of movement or joint damage). The present sequence
 CC represents human megakaryocyte stimulating factor (MSF, CACP protein).
 CC Note: This sequence is not given in its entirety in figure 4 of the
 CC specification, although a GenBank accession number was given. This
 CC sequence was therefore obtained from GenBank (U70316).

CC Sequence 1404 AA;

Query Match 100.0%; Score 5011; DB 22; Length 1404;
 Best Local Similarity 100.0%; Pred. No. 1,6e-273;
 Matches 941; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKDNKKNTKKKPPPVNDAGSLNDGDFKVTPTSTTQHNKYSTSKITTAAPIN 60
 DB 200 VKDNKKNTKKKPPPVNDAGSLNDGDFKVTPTSTTQHNKYSTSKITTAAPIN 259
 QY 61 PRPSLPNSDTSKETSLSLVNKEETVETKETTNNKOTSDGKEKTTSAKQSTKTSK 120
 DB 260 PRPSLPNSDTSKETSLSLVNKEETVETKETTNNKOTSDGKEKTTSAKQSTKTSK 319
 QY 121 DLAPTSKVLAPPTKALETTTGPAITTPKEPTPTTPKEPASTPREPTTITKSAPTPK 180
 DB 320 DLAPTSKVLAPPTKALETTTGPAITTPKEPTPTTPKEPASTPREPTTITKSAPTPK 379
 QY 181 EPAPTTTASAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 240
 DB 380 EPAPTTTASAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 439
 QY 241 PKAPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEP 300
 DB 440 PKAPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEP 499
 QY 301 PTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKE 360
 DB 500 PTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKE 559
 QY 361 PAPTTPKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPT 420
 DB 560 PAPTTPKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPT 619
 QY 421 TPKKLTPTTPEKLAFTTPEKAPATTPEELAPTTPEEPPTTPEEPAPTTKAAAPNTPE 480
 DB 620 TPKKLTPTTPEKLAFTTPEKAPATTPEELAPTTPEEPPTTPEEPAPTTKAAAPNTPE 679
 QY 481 PAPTTPKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPT 540
 DB 680 PAPTTPKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPT 739
 QY 541 EPTSTTSKAPATTPKGTATTPKEPAPTTPKAPATTPKGTATTPKEPAPTTPKAP 600
 DB 740 EPTSTTSKAPATTPKGTATTPKEPAPTTPKAPATTPKGTATTPKEPAPTTPKAP 739
 QY 601 KELAPTTKGPSTTSKAPATTPKGTATTPKEPAPTTPKAPATTPKGTATTPKEPAPT 660
 DB 800 KELAPTTKGPSTTSKAPATTPKGTATTPKEPAPTTPKAPATTPKGTATTPKEPAPT 859
 QY 661 PTTTKEPTTTHKSPDSTPELSAETPKALENSPKEGVTTTTPATPEMTTAKDKT 720
 DB 860 PTTTKEPTTTHKSPDSTPELSAETPKALENSPKEGVTTTTPATPEMTTAKDKT 919
 QY 721 TEBDLATTPETTTAAKPKMKETATTTTEKTTESKITATTTQVSTTTQDTTPFKITLKT 780
 DB 920 TEBDLATTPETTTAAKPKMKETATTTTEKTTESKITATTTQVSTTTQDTTPFKITLKT 979

QY 761 TLAPKVTYTKKNTTTEIMNKPPEETAKPRDRATNSKATTPKPKPTAKKPTSTKPKPT 840
 DB 980 TLAPKVTYTKKNTTTEIMNKPPEETAKPRDRATNSKATTPKPKPTAKKPTSTKPKPT 1039
 QY 841 MPRAKPKPTTTPPKMTSTMPBLNPTSRIABAMLOTTTRPNOTPNSKLVEVNPSEDA 900
 DB 1040 MPRAKPKPTTTPPKMTSTMPBLNPTSRIABAMLOTTTRPNOTPNSKLVEVNPSEDA 1099
 QY 901 AEGTPEHMLRPVFMPEVTPDMDYLPRVPMOGIITINPMLS 941
 DB 1100 AEGTPEHMLRPVFMPEVTPDMDYLPRVPMOGIITINPMLS 1140

RESULT 4
 AAB29773
 ID AAB29773 standard; Protein: 1404 AA.
 AC AAB29773;
 DT 28-FEB-2001 (first entry)
 DE Human megakaryocyte stimulating factor (MSF), SEQ ID NO:1.
 XX
 XX
 XX
 KW Human MSF, megakaryocyte stimulating factor; tribonectin;
 KW alternative splicing; joint boundary lubricant; O-linked oligosaccharide;
 KW osteoarthritis; tribosupplementation; tissue adhesion inhibition;
 KW friction coefficient reduction; gene therapy; antiarthritic;
 KW osteopathic.
 XX
 XX
 OS Homo sapiens.
 PN WO200064930-A2.
 XX
 XX
 PD 02-NOV-2000.
 XX
 XX
 PF 24-APR-2000; 2000MO-US10953.
 XX
 PR 23-APR-1999; 99US-0298970.
 XX
 XX
 PA (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.
 XX
 PI Jay GD;
 XX
 DR WPI: 2001-024673/03.
 DR N-FSDB; AAC81498.
 XX
 PT Novel tribonectin polypeptide useful as lubricant for treating
 PS osteoarthritis, comprises O-linked lubricating moiety
 PS
 PS Claim 3; Page 7; 47pp; English.

CC The invention relates to a human tribonectin which is a product of
 CC alternative splicing of the human MSF (megakaryocyte stimulating factor)
 CC gene. The tribonectin has at least one O-linked oligosaccharide
 CC lubricating moiety and has a polypeptide sequence comprising 1-76
 CC repeats of a motif having at least 50% identity to the sequence KEPAPTT
 CC (AAB29774). The invention also relates to a nucleic acid encoding a
 CC human MSF-derived tribonectin; a biocompatible composition comprising a
 CC human tribonectin for inhibiting tissue adhesion formation; and a method
 CC of diagnosing osteoarthritis or a predisposition to osteoarthritis by
 CC measuring the amount of MSF or its fragment in a biological sample of a
 CC mammal, wherein an increased amount of MSF compared to a control
 CC indicates the presence of or predisposition to developing
 CC osteoarthritis. The tribonectin and DNA encoding it are useful in the
 CC treatment of osteoarthritis, where they may be used for lubricating
 CC mammalian joints, such as articulating joints of humans, dogs or horses.
 CC The tribonectin, when formulated as a membrane, foam, gel or fibre, is
 CC useful for inhibiting adhesion between two surfaces such as the injured
 CC tissues of a mammal, where the injury is caused by a surgical incision
 CC or trauma, or an artificial device e.g., an orthopaedic implant. In
 CC particular, one of the surfaces is pericardial tissue. DNA encoding a
 CC tribonectin may be used in gene therapy. The present sequence represents

PT immune deficiencies, cancer, exposure to radiation or drugs,
 PT bacterial and viral infections, etc.

PS Claim 1, 2 and 3; Fig 1: 87pp; English.

XX The sequence given is a full length translation from the megakaryocyte
 CC stimulating factor (MSF) precursor. The sequence covered by exons II,
 CC III and IV encodes megakaryocyte stimulating factor (MSF). This
 CC sequence is modified by the addition of an N-terminal sequence encoding
 CC a secretory leader, an initiating methionine proceeding exon II and a
 CC terminating codon following exon IV. The cDNA sequence given contains
 CC sequences derived from human megakaryocyte colony stimulating factor
 CC (meg-CSF). Exon I contains the initiating methionine, and encodes a
 CC classical mammalian protein secretion signal sequence. The sequence
 CC encoding the original meg-CSF includes exons II-IV and is thought to
 CC terminate in the region between amino acid residues 134 - 147. The
 CC primary transcript of this gene may be cleaved in different ways to
 CC yield a family of mRNA's each encoding a different MSF protein. Exons
 CC V and VI are thought to be related to the activity of the factor and
 CC are also implicated in the stability, folding and processing of the
 CC molecule. These exons are also thought to play a role in the observed
 CC synergy of MSF with other cytokines. Exons V - XII are believed to be
 CC implicated in the processing or folding of the appropriate structure of
 CC the resulting factor i.e. one or more of these exons may contain
 CC sequences which direct proteolytic cleavage, adhesion, organization of
 CC the cellular matrix or extracellular matrix processing. Both naturally
 CC occurring and non-naturally occurring MSF's may be characterised by
 CC various combinations of alternatively spliced exons from this sequence,
 CC with the exons spliced together in differing orders to form different
 CC members of the MSF family.

XX Sequence 1404 AA:

SQ Query Match 100.0%; Score 5011; DB 13; Length 1404;
 Best Local Similarity 100.0%; Pred. No. 1.6e-273;
 Matches 941; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKDNKKNRTRKKKPPKPPVVDAGSGLDNGDFKVTTPDTSTGHNKYSTSPKITTAKPIN 60
 DB 200 VKDNKKNRTRKKKPPKPPVVDAGSGLDNGDFKVTTPDTSTGHNKYSTSPKITTAKPIN 259
 QY 61 PRPSLPNSTSTSTLVNKEETVETKETTNNKORTSGKEKETSAAKESOSTEKSATK 120
 DB 260 PRPSLPNSTSTSTLVNKEETVETKETTNNKORTSGKEKETSAAKESOSTEKSATK 319
 QY 121 DLAPTSKVLAKPPPKAEETTKGSPALTTTPKEPTTPKEPASTPKEPTTTIKSAPTPK 180
 DB 320 DLAPTSKVLAKPPPKAEETTKGSPALTTTPKEPTTPKEPASTPKEPTTTIKSAPTPK 379
 QY 181 EPAPPTTKSAPPTTPKEPAPPTTKKEPAPPTTKKEPAPPTTKSAPPTTKKEPAPT 240
 DB 380 EPAPPTTKSAPPTTPKEPAPPTTKKEPAPPTTKKEPAPPTTKSAPPTTKKEPAPT 439
 QY 241 PKKRAPTPPKRAPPTTPKEPAPPTTPKEPAPPTTKKEPAPPTTKKRAPTPPKRA 300
 DB 440 PKKRAPTPPKRAPPTTPKEPAPPTTPKEPAPPTTKKEPAPPTTKKRAPTPPKRA 499
 QY 301 PTPKPEAPPTTKKEPAPPTTPKEPAPPTTKSAPPTTKSAPPTTKKEPAPPTTKKE 360
 DB 500 PTPKPEAPPTTKKEPAPPTTPKEPAPPTTKSAPPTTKSAPPTTKKEPAPPTTKKE 559
 QY 361 PAPPTTPKRAPPTTPKRAPPTTPKRAPPTTPKRAPPTTKKRAPTPPKRAPPTTKRA 420
 DB 560 PAPPTTPKRAPPTTPKRAPPTTPKRAPPTTPKRAPPTTKKRAPTPPKRAPPTTKRA 619
 QY 421 TPKKLTPTPEKLAAPTPPEKAPPTPEELAPPTPEEPPTTPPEEPAPTPPKAAAPNPK 480
 DB 620 TPKKLTPTPEKLAAPTPPEKAPPTPEELAPPTPEEPPTTPPEEPAPTPPKAAAPNPK 679
 QY 481 PAPPTTPKRAPPTTPKRAPPTTPKRAPPTTPKRAPPTTKKRAPTPPKRAPPTTKRA 540
 DB 680 PAPPTTPKRAPPTTPKRAPPTTPKRAPPTTPKRAPPTTKKRAPTPPKRAPPTTKRA 739

QY 541 EPTSTSDKAPAPTPPKGTAATTPKKEPAPPTTKKEPAPPTTKGTAATTLKEPAPTPPKRAP 600
 DB 740 EPTSTSDKAPAPTPPKGTAATTPKKEPAPPTTKKEPAPPTTKGTAATTLKEPAPTPPKRAP 799
 QY 601 KELAPPTTKGPTSTSDKRAPPTPKETAPTPKEPAPPTTKKRAPPTTPKEPTTSEVST 660
 DB 800 KELAPPTTKGPTSTSDKRAPPTPKETAPTPKEPAPPTTKKRAPPTTPKEPTTSEVST 859
 QY 661 PPTTKEPTTKHSPPDESPTPELSAEPTPKALNSPKPEGPTTKTPAANKPEMTTAKOKT 720
 DB 860 PPTTKEPTTKHSPPDESPTPELSAEPTPKALNSPKPEGPTTKTPAANKPEMTTAKOKT 919
 QY 721 TERDRTPEPTTAAPKMKKEPATTTEKTESKITATTQVSTTODTPPKITTLKTT 780
 DB 920 TERDRTPEPTTAAPKMKKEPATTTEKTESKITATTQVSTTODTPPKITTLKTT 979
 QY 781 TLAPKVTTKKTIITTEINMKPEETAKPKDRATNSKATTPKQOKPTKAPKSTTKKPT 840
 DB 980 TLAPKVTTKKTIITTEINMKPEETAKPKDRATNSKATTPKQOKPTKAPKSTTKKPT 1039
 QY 841 MPVVRKPKTTTPPKKMTSMPELNPSTRIAEAMLOTTTPNCTPNSKILVEVNPKEPDAG 900
 DB 1040 MPVVRKPKTTTPPKKMTSMPELNPSTRIAEAMLOTTTPNCTPNSKILVEVNPKEPDAG 1099
 QY 901 AEGETPHMLRPHEVMEPEVTPOMDYLRVYNOCIIINPMLS 941
 DB 1100 AEGETPHMLRPHEVMEPEVTPOMDYLRVYNOCIIINPMLS 1140

RESULT 3

AAB60568
 ID AAB60568 standard; Protein: 1404 AA.

AC AAB60568;

DT 27-APR-2001 (first entry)

XX Human megakaryocyte stimulating factor (MSF, CACP).

XX Human: CACP protein; campodactylly-arthropathy-coxa vara-pericarditis;

KW MSF; megakaryocyte stimulating factor; synovial lubricant;

KW chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;

XX antiarthritic.

OS Homo sapiens.

XX WO200107068-A1.

PD 01-FEB-2001.

PF 21-JUL-2000; 2000MO-US20002.

XX 23-JUL-1999; 99US-0145328.

PR 19-JUL-2000; 2000US-0145328.

XX (UYCA-) UNIV CASE WESTERN RESERVE.

PI Warman ML;

DR WPL; 2001-182721/18.

XX New composition comprising the campodactylly-arthropathy-coxa
 PT vara-pericarditis protein in combination with an anesthetic, useful for
 PT treating osteoarthritis, or as lubricants of tissue and joints

PS Example 1; Page -: 34pp; English.

XX The invention relates to a method of treating osteoarthritis via the
 CC administration of a composition comprising the campodactylly-arthropathy-
 CC coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.
 CC The composition may further comprise a local anesthetic. The composition
 CC of the invention may be administered via intra-articular or intravenous

DR N-PSDB: AAH98981.
 XX Isolated polypeptide for treatment of diseases, diagnostics, raising
 PT antibodies and research use -
 XX
 XX
 PS Claim 20: Page 1198-1201: 1275pp; English.
 CC The present invention provides the protein and coding sequences of novel
 CC proteins from a variety of organisms, including human, dog, cat, horse,
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
 CC from the organism of interest. They can be used in diagnostics,
 CC forensics, gene mapping, identification of mutations, to assess
 CC biodiversity and for nutritional purposes. The present sequence is a
 CC protein of the invention.
 XX
 XX Sequence 1299 AA;

Query Match 100.0%; Score 5011; DB 22; Length 1299;
 Best Local Similarity 100.0%; Pred. No. 1.5e-273;
 Matches 941; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VADNKKNRTPKPPVNDAGSLGNDGPKYTPPTSTTQHNKYSTSKITTAAPIN 60
 DB 200 vkdnknkttckkppkpvndagsgldngdtkvtlptdstctqhnkystskitlakpin 259
 QY 61 PRPSLRPNSDTSKETSILVNKEETVETKETTNNKQSTDGKEKTSKETSKEKTSK 120
 DB 260 prpslprnsdtsketsilvnkettvetkettnnkqstdgketsketsketsktsak 319
 QY 121 DLAPTSKVLAPPTKAEETTGAPALTTPKEPTPTTPKEPASTTPKEPTTIKSAPTTPK 180
 DB 320 dlaptskvlapptkaeettgpaltpkepttpkepaalttkpepaalttkpepttk 379
 QY 181 EPAPTTKASAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTT 240
 DB 380 epapttksapttkepapttkepapttkepapttkepapttkepapttkepapttke 439
 QY 241 PKKAPPTPKKAPPTPKKAPPTPKKAPPTPKKAPPTPKKAPPTPKKAPPTPKKAP 300
 DB 440 pkkapptpkkapptpkkapptpkkapptpkkapptpkkapptpkkapptpkkap 499
 QY 301 PPTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKE 360
 DB 500 ptpkepapttkepapttkepapttkepapttkepapttkepapttkepapttke 559
 QY 361 PAPPTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAP 420
 DB 560 papptkepapttkepapttkepapttkepapttkepapttkepapttkepaptt 619
 QY 421 TPKLTPTTPEKLAAPTTPEKAPPTPEELAPTTPEEPTPTTPEEAPTTKAAAPNTPK 480
 DB 620 tpkltpttpeklaapttpekaptpeelapttpeeptpttpeeapttkaapntpk 679
 QY 481 PAPPTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAP 540
 DB 680 papptkepapttkepapttkepapttkepapttkepapttkepapttkepaptt 739
 QY 541 EPTSTTSKAPPTPKGAPTTPKKAPPTPKKAPPTPKKAPPTPKKAPPTPKKAP 600
 DB 740 eptsttskapptpkgapptpkkapptpkkapptpkkapptpkkapptpkkap 699
 QY 601 KELAPTTTGKPTSTSDKAPPTPKETAPTTPKKAPPTPKKAPPTPKKAPPTPKK 660
 DB 800 kelaptttgkptstsdkapptpketaptpkkapptpkkapptpkkapptpkkap 759
 QY 661 PTTTKSPPTTIHKSPTSDELSEAPTPKALENPKKPGVPTTTPATPEMTTAKDXT 720
 DB 860 ptttksppttihksptsdelseaptpkalenpkpgvptttpatpemttaakdxt 919
 QY 721 TERDLTTTETTTAAKMKKETAATTEKTESKIATTTQVSTTTQDTPPKITTLKTT 780

DB 920 terdlrttpeettaapkmktetattlekteskiltattgvtsetgdltpfkiltlkt 979
 QY 791 TLAPKVVTTTKKTTTTEITMNRPEETAKPKDRATNSKATTPKOKPTKAPKPKTKKPK 840
 DB 980 tlapkvtttkktttiteitmnrpeetakpkdratskattpkpkpkpkpkpkpkpkpk 1039
 QY 841 MPVRKPKKTTTPPKKMTSTMBELNPTSRIAPAMQTTTRPNQTPNSKLVEVNPSEDA 900
 DB 1040 mpvrkpkktttppkkmtstmbelnptsriaeamqgttrtpnqtpnslvevnpse 1099
 QY 901 AEGETPHMLLRPHVFMPEVTPDMDYLPRVPMQGIITPMIS 941
 DB 1100 aegetphmllrphvfmpevtpdmdylprvpmqgiitpmis 1140

RESULT 2
 AAR26049
 ID AAR26049 standard; Protein: 1404 AA.
 XX
 AC AAR26049;
 XX
 DT 02-FEB-1993 (first entry)
 XX
 XX MSF precursor.

KW Megakaryocyte colony stimulating factor; secretion signal; meg-CSF;
 KW stability; proteolytic cleavage; adhesion; alternative splicing.
 OS Synthetic.

FT Key Location/Qualifiers
 FT Region 1..26 /label= Exon_I
 FT Region 26..67 /label= Exon_II
 FT Region 67..107 /label= Exon_III
 FT Region 107..157 /label= Exon_IV
 FT Region 157..200 /label= Exon_V
 FT Region 200..1141 /label= Exon_VI
 FT Region 1141..1166 /label= Exon_VII
 FT Region 1166..1212 /label= Exon_VIII
 FT Region 1213..1266 /label= Exon_IX
 FT Region 1266..1331 /label= Exon_X
 FT Region 1331..1373 /label= Exon_XI
 FT Region 1373..1404 /label= Exon_XII

FT WO9213075-A.
 FT 06-AUG-1992.
 FT 17-JAN-1992: 92WO-US00433.
 FT 18-JAN-1991: 91US-0643502.
 FT 10-SEP-1991: 91US-0757022.
 FT (GENY) GENETICS INST INC.

PI Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K;
 DR WPI: 1992-284660/34.
 DR N-PSDB: AAQ27223.
 PT New human mega-karyocyte stimulating factors - for treating

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:13:29 ; Search time 107.17 Seconds
(without alignments)
650.396 Million cell updates/sec

Title: US-09-556-246-1_COPY_200_1140
5011
Perfect score: 1 VKDKKKRTRKKRPKPPV.....DMDYLPVPNGIINMELS 941
Sequence:

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: /SIDSR/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDSR/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDSR/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDSR/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDSR/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDSR/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDSR/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDSR/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDSR/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDSR/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDSR/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDSR/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDSR/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDSR/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDSR/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDSR/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDSR/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDSR/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDSR/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDSR/gcgdata/geneseq/geneseq/AA2000.DAT.*
21: /SIDSR/gcgdata/geneseq/geneseq/AA2001.DAT.*
22: /SIDSR/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5011	100.0	1299	22	Human EST encoded
2	5011	100.0	1404	13	MSF precursor. Sy
3	5011	100.0	1404	22	Human megakaryocyt
4	5011	100.0	1404	22	Human megakaryocyt
5	2812.5	56.1	902	22	Human MSF-derived
6	1357.5	27.1	452	16	Human megakaryocyt
7	1168	23.3	5179	22	C899p predicted am
8	981	19.6	763	21	Arabidopsis thaliana
9	942	18.8	1664	21	C. thermocellum OI
10	771.5	15.4	1325	22	Peptide #2327 enco
11	715	14.3	763	18	Mycobacterium tube

12	625.5	12.5	4412	21	AAV53666
13	612	12.2	572	18	AAW31855
14	544	10.9	844	7	AAV60570
15	542	10.8	807	21	AAV54467
16	520.5	10.4	788	21	AAV54466
17	506.5	10.1	744	9	AAV82975
18	496.5	9.9	2971	21	AAV41231
19	493	9.8	1837	21	AAV11726
20	489.5	9.8	2972	22	AAV50365
21	489.5	9.8	3118	22	AAV50362
22	488.5	9.7	826	13	AAV26042
23	488	9.7	617	22	AAV14858
24	488	9.7	617	22	AAV04187
25	488	9.7	957	21	AAV59288
26	488	9.7	957	21	AAV52813
27	485	9.7	1127	22	AAV95541
28	476.5	9.5	652	9	AAV82974
29	467.5	9.3	511	22	AAV14883
30	467.5	9.3	511	22	AAV27312
31	467.5	9.3	511	22	AAV02607
32	466	9.3	1721	19	AAV48299
33	463.5	9.2	1721	21	AAV11727
34	450.5	9.0	378	12	AAV14160
35	446.5	8.9	378	12	AAV14162
36	446.5	8.9	750	20	AAV05477
37	445.5	8.9	472	22	AAV60569
38	443	8.8	2870	21	AAV95559
39	443	8.8	3178	21	AAV95556
40	442.5	8.8	751	16	AAV80839
41	425.5	8.5	910	22	AAV83007
42	424.5	8.5	1012	20	AAV17406
43	424.5	8.5	1125	21	AAV22934
44	424.5	8.5	1125	21	AAV79637
45	424.5	8.5	1610	21	AAV22870

ALIGNMENTS

RESULT 1	
ID	AAW24322 standard; Protein; 1299 AA.
XX	AAW24322:
XX	
DF	12-OCT-2001 (first entry)
XX	
DE	Human EST encoded protein SEQ ID NO: 1847.
XX	
KW	Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW	tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW	diagnostics; forensic test; gene mapping; genetic disorder;
KW	biodiversity; gene therapy; nutrition.
XX	
OS	Homo sapiens.
XX	
PN	W020015447-A2.
XX	
PD	02-AUG-2001.
XX	
PF	25-JAN-2001; 2001MO-US02687.
XX	
PR	25-JAN-2000; 2000US-0491404.
PR	17-JUL-2000; 2000US-0617746.
PR	03-AUG-2000; 2000US-0631451.
PR	15-SEP-2000; 2000US-0663870.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI	Gao Y, Drmanac RA, Zhang J, Werhman T;
XX	
DR	WPI; 2001-476164/51.

Sequence q1/101742
Mycobacterium tube
Sequence of the fa
Amin acid sequenc
Amino acid sequenc
Bioadhesive precu
Human OREF ORF95
Cryptosporidium pa
Human SRCAP. Homo
Human SRCAP. Homo
P. yoelli SSP2 ant
Peptide #2892 enco
Peptide #2869 enco
Human MUC11 polype
C900P predicted am
Human protein sequ
Bioadhesive precu
Peptide #1317 enco
Peptide #1349 enco
Peptide #1289 enco
Cryptosporidium pa
Portion of Cryptos
ppp 378. Trilicium
ppp encoded by clo
C. albicans Rbt1 p
Bovine MSF ortholo
Caenorhabditis ele
Japanese sea musse
S. epidermidis ope
Human atrophin-1 r
Mouse microtubule-
Microtubule associ
EVFP-DEVD-MAF4-EBF

XX WPI: 2001-476286/51.
 XX Novel single exon nucleic acid probe used to measuring gene expression
 PT in a human breast -
 XX
 PS Claim 27; SEQ ID No 12385; 322pp; English.
 XX
 CC The present invention relates to novel single exon nucleic acid probes
 CC (see AA100010-AA110067). The present sequence is a peptide encoded by one
 CC such as AA100010-AA110067). The present sequence is a peptide encoded by one
 CC a human breast sample, where the probe hybridises at high stringency to a
 CC nucleic acid expressed in the human breast. The probes are useful for
 CC predicting, diagnosing, grading, staging, monitoring and prognosing
 CC diseases of the human breast, particularly those diseases with polygenic
 CC aetiology. The diseases include: breast cancer, disorders of development,
 CC inflammatory diseases of the breast, fibrocystic changes, proliferative
 CC breast disease and non-carcinoma tumours.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.
 CC
 XX Sequence 1325 AA:

Query Match 15.0%; Score 771.5; DB 22; Length 1325;
 Best Local Similarity 26.0%; Pred. No. 9.1e-36;
 Matches 302; Conservative 97; Mismatches 466; Indels 295; Gaps 45;

QY 7 NRRK-KRPKPP-PPVDEAGSGLDNGDFKVTTPDTSTQO-----HNKVSSTPKITP 55
 Db 60 ntrrhkprdkpckgnsktdihk--stidhheaptsseensngkdpmlnrgsvdpdst 118
 QY 56 A-----KPINPRSLPN--SDTSKET-----SLTVNKETVENKET 90
 Db 119 tlhkasagkhilpapkskincrstgkstvtrksdktgrpleksmldktsksht 178
 QY 91 TTT-----NKOTSDGKETTSAKET-----QSIETK-SAKDLAPTSKVLAKPT 133
 Db 179 ttfmshgmsqtkstksfpekitaasktyktqtpseesektedsttvasdkllklt 238
 QY 134 PKAETTTKGALP-----TPKEPPT--TPKEPASTTPKEPPT 170
 Db 239 knqgetisanelqslaepthegrtanentlpsaepnterantentlpsaepth 298
 QY 171 TTKSA--PTTPKEPAP-----TTKSAPTTPKAPATTKE--PAPTTKEPAPT 216
 Db 299 rextanentapfpaqptenremtanentlfpaeptegertanentlpsaepthger 358
 QY 217 TTKEPAPTTKSAPTTPKEPAPTTPKKAPTTTPKEPAPTTPKEPPT--TPKEPAPT 271
 Db 359 tane-----ntpsaepthegertpandktlssaeptegertplanentlpsaep 414
 QY 272 -TKEPAPTTKAPAPTTAPKAPATTTPKEPAPTTPKEPA--PTTKKPSPTTPKEPA-- 324
 Db 415 enertanentlpsaepnterantentlpsaepnterantentlpsaepnterant 474
 QY 325 -----PTTKSA-PTTKKEPAPTTKSAPTTPKEPPTTKE-----PAPTPK 367
 Db 475 qtrpfanektssaeptegertplanen--tlpsaepnterantentlpsaepth 532
 QY 368 EPAPTTKAPAPTTTPKEPAPT-----TPKEPAPTTKAPAPTTAPKAPATTTPK 422
 Db 533 engdrplanektlpsaepnterantentlpsaepnterantentlpsaepnterant 589
 QY 423 KKLPTTPKELAPTPPKAPAPTTPEELAPTTPEEPPTTPKEPAPTTPKAAPNT--PK 479
 Db 590 speptenrertanentlpsaepnterantentlfpaeptenrertanentlpsa 649
 QY 480 EPAPTTKAPAPTTTPKEPAPTTPKAPATTTPKGAAPTTLKAPAP-----TPKKAPKE 533
 Db 650 eptengqtrpfanektssaeptegertplanentlpsaepnterantentlpsa 709

QY 534 LAPTTKEPSTTSDKAPATTPKCTAPTTPKAPATTTPKEPAPT-----TPKGTAPTTLK 588
 Db 710 aeptenrertanentlpsaepnterantentlpsaepnterantentlpsaepnterant 769
 QY 589 EPAPT-----TPKKA-PKE-----LAPTTKGPSTTSDKAP 621
 Db 770 saeptehaertplanentlpsaepnterantentlpsaepnterantentlpsaepnterant 829
 QY 622 TTPKET-----APTTKAPATTTPKAPATTTPKEPAPTTPKAPATTTPKAPATTTPK 670
 Db 830 aeptenrertanentlpsaepnterantentlpsaepnterantentlpsaepnterant 889
 QY 671 HKSPDE-----STPELSAEPKPK-----ALENSPKEPVPT-----TKTPA 706
 Db 890 saeptehaertplanentlpsaepnterantentlpsaepnterantentlpsaepnterant 948
 QY 707 ATPD-----EMTTAKDVTTERDLKTTPETTTAPKMTKE-----TATTEKYTE 751
 Db 949 saeptengertplanentlpsaepnterantentlpsaepnterantentlpsaepnterant 1008
 QY 752 SKITATP-TQVSTTTQDTPPKI-----TTL-----KTTTLAPK 785
 Db 1009 spakpleheemtpsaenentlpsaepnterantentlpsaepnterantentlpsaepnterant 1068
 QY 786 VTTT-----KTIITTEIMNKPETAKPKDRATNSKATTPPKOKP----- 825
 Db 1069 tlpslakphtegertlpsaepnterantentlpsaepnterantentlpsaepnterant 1128
 QY 826 -----TKADKPKPTSTKPKKTPMPKPKP-----KTPPTPKMTSTMBEL 863
 Db 1129 ekmtqvtelstehepkektstte-ktlrtpexklyseklctkgnpvpkeptenlgn 1187
 QY 864 NPTSRINAMLOTTPRPNOT 883
 Db 1188 tlletlikavkstenpekt 1207

RESULT 11
 AAW31852
 ID AAW31852 standard; Protein; 763 AA.
 XX
 AC AAW31852;
 XX
 DT 27-APR-1998 (first entry)
 XX
 DE Mycobacterium tuberculosis 74 kda protein.
 XX
 KW Tuberculosis; mycobacteria; infection; diagnosis;
 KM antimycobacterial; antibiotic; vaccine.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN W09741252-A2.
 XX
 PD 06-NOV-1997.
 XX
 PF 18-APR-1997; 97WO-EP01973.
 XX
 PK 29-APR-1996; 96DE-4017184.
 XX
 PA (GBFB) GBF GES BIOTECH FORSCHUNG GMBH.
 XX
 PI Espitia C, Honisch C, Moreno C, Singh M;
 XX
 DR WPI: 1997-549750/50.
 XX
 DR N-PSDB; AAT93610.
 XX
 PT New DNA and related proteins or RNA derived from M. tuberculosis
 PT used for diagnosis of mycobacterial infections, monitoring
 PT vaccination and development of anti-mycobacterial agents
 XX
 PS Claim 5; Fig 13; 55pp; English.

```

FT      Domain      409..565
FT      /note="cohesin type II domain"
FT      Domain      607..763
FT      /note="cohesin type II domain"
XX
XX      FR2748479-A1.
XX
XX      14-NOV-1997.
XX
XX      10-MAY-1996; 96FR-0005854.
XX
XX      10-MAY-1996; 96FR-0005854.
XX
XX      (INSP ) INST PASTEUR.
XX
XX      Beguin P, Leibovitz E;
XX
XX      WPI: 1998-011569/02.
XX      N-PSDB; AAT86623.
XX
XX      Cellulase proteins with cohesin or dockerin type II domains - useful
XX      for potentiating the activity of multiprotein enzyme complexes
XX
XX      Claim 7; page 31-39; 60pp: French.
XX
XX      Multimeric protein, especially enzymatic, complexes are held together
XX      by protein-protein interactions between domains designated dockerins
XX      and cohesins, which are found on the catalytic and scaffold subunits
XX      respectively. An example of such a complex is the cellulose degrading
XX      protein complex from Clostridium thermocellum, known as the cellulosome.
XX      This complex comprises around 15 proteins including endoglucanases,
XX      cellobiohydrolases, hemicellulases, e.g. xylnases or lichenanases, which
XX      interact with a central "scaffold" protein designated the cellulosome
XX      interacting protein (CIPA; see AAM43108). The catalytic subunits
XX      interact with the CIP subunit via conserved 23 amino acid dockerin
XX      domains. CIP has been shown to contain 9 copies of a cohesin domain.
XX      The invention relates to the isolation of proteins binding to a novel
XX      dockerin type domain found in the C-terminal portion of CIP. The new
XX      domain is designated a type II dockerin domain (as compared to the type
XX      I domain found on the catalytic subunits of the cellulosome). The type
XX      II dockerin domain has some sequence similarity to the type I dockerins
XX      but is unable to bind type I cohesin domains.
XX      The sequence presented here is an example of a protein which binds
XX      the novel type II dockerin domain and is the product of the OIPB gene.
XX      The protein contains 4 type II cohesin domains in the N-terminal portion
XX      of which the first domain (amino acid residues 28-192) is thought to
XX      bind CIPA. The novel type II dockerin and cohesin domains can be used
XX      in complexes, especially enzyme complexes, to potentiate their catalytic
XX      actions in a synergistic manner.
XX
XX      Sequence 1664 AA:
XX
XX      Query Match      18.3%; Score 942; DB 19; Length 1664;
XX      Best Local Similarity 33.0%; Pred. No. 3e-45;
XX      Matches 276; Conservative 83; Mismatches 312; Indels 166; Gaps 43;
XX
XX      QY      128 VLAKPP-KAETTTGCPALTPPEPTTPPKPEASTTPPEPTTITKSAPTTPKEPAPTT 186
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      Db      758 VYIYPAPILKaaadepIptldtspdepas-----depus---depltsdepltp 804
XX
XX      QY      187 TKSAPTTPKEPAPTTTPKEPA-----PTTTPKEPAPTTTSAPTTPKEPAPTT 242
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      Db      805 eptpsctpeepIptldtspdepIptldtspdepIptldtspdepIptldtspdepIptldtsp 864
XX
XX      QY      243 KAPATTPKEPAPTTTPKEPTTPTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 301
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      Db      865 adeplpsdeplps--deplps--deplp--sdeplpsdeplpsdeplpsdeplpsdeplp 919
XX
XX      QY      302 TTPKKA--PTTTPKEPSPTTPKEPAPTTTPKS-APTTPKEPAPTTTPKSAPTTPKEPSPTT 357
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      Db      920 sdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplp 968

```

```

QY      358 TKEPAPTTPKEPAPTTTPKKAAPTTPKKEPAPTTTPKKEPAPTTTPKKAAPTTPKKE 416
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      Db      969 sdeplps--deplps-----deplpsdeplpsdeplpsdeplpsdeplpsdeplpsd--- 1019
XX
XX      QY      417 TAPTPPKLTPTPEKLAPTTPKEKAPPTPEELAPTTPEEPTPTTPKEEPAPTTPKKAAP 475
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      Db      1020 -----eplpsde--eplpsdeplpsd-----eplpsdeplpsdeplpsdeplp 1064
XX
XX      QY      476 NTPKEPAPTTPKEPA--PTTPKKEPAAPTTPKEPAAPTTPKKAAPTTPKKAAPTTPKKAAP 532
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      Db      1065 tpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplp 1122
XX
XX      QY      533 ELAPTTPKEPT-STISDKAPAPTTPKGAAPTTPKEPAAPTTPKKEPAAPTTPKKAAPTTPKKEPA 591
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      Db      1123 d-eplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplp 1175
XX
XX      QY      592 PT-TPKKAPAKELAPTTTPKGPSTTSDDKPAAPT--TPKKEPAAPTTPKKEPAAPTTPKKAAPTTP 646
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      Db      1176 psetpseep-----lptldtspdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplp 1229
XX
XX      QY      647 TPETTPPTTSEVSTPTTTPKEPTTHKSPDESTBELSAEPTPKALENSPKRGVPTTKTPA 706
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      Db      1230 dtpsddeplpsd--eplpsdeplp-----psdeplp--sdeplp--lptldtsp 1277
XX
XX      QY      707 ATKPEMTTAKDKTTEREDLRTTPETTTAAPKMKETATTTTEKTESKITATTTQVSTTT 766
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      Db      1278 eplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplps 1337
XX
XX      QY      767 QDTTPPKITTLKTTTLAPKVTTTKKTTTTEIMNKPEETAKPKDRATNSKATPKPKP-KP 825
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      Db      1338 deplpsdeplp-----psdeplpsdeplpsdeplpsdeplpsdeplpsdeplp 1372
XX
XX      QY      826 TKAPKKEPTSTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTK 870
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      Db      1373 setpseplp---lptldtspstpsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsg 1424
XX
XX      QY      871 EAMLQTTTPNCPNKSILVEVNPKSSEDAGAEGETHMLLRPVHEPTTPMDIYTP 927
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      Db      1425 -----kptseplpdele-eplpsdvpgaIygehrayIry-----pdgsftr 1465
XX
XX      RESULT 10
XX      AAM03645
XX      ID      AAM03645 standard; Protein: 1325 AA.
XX      XX
XX      AC      AAM03645;
XX      XX
XX      DE      09-OCT-2001 (first entry)
XX      XX
XX      DE      Peptide #2327 encoded by probe for measuring breast gene expression.
XX      XX
XX      KW      Probe; human; breast disease; breast cancer; development disorder;
XX      XX      inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX      XX
XX      OS      Homo sapiens.
XX      XX
XX      PN      WO200157270-A2.
XX      XX
XX      PD      09-AUG-2001.
XX      XX
XX      PF      29-JAN-2001; 2001WO-US00661.
XX      XX
XX      PR      04-FEB-2000; 2000US-0180312.
XX      PR      26-MAY-2000; 2000US-0207456.
XX      PR      30-JUN-2000; 2000US-0608408.
XX      PR      03-AUG-2000; 2000US-0632366.
XX      PR      21-SEP-2000; 2000US-0234687.
XX      PR      27-SEP-2000; 2000US-0236359.
XX      PR      04-OCT-2000; 2000GB-0024263.
XX      XX
XX      PA      (MOLE-) MOLECULAR DYNAMICS INC.
XX      XX
XX      PI      Penn SG, Hanzel DK, Chen W, Rank DR;

```

```
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148684.
PR 13-AUG-1999; 99US-0149368.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151338.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 18-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
```

```
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 19.0%; Score 981; DB 21; Length 763;
Best Local Similarity 38.5%; Pred. No. 8.9e-48;
Matches 226; Conservative 24; Mismatches 293; Indels 44; Gaps 11;

```
OY 137 ETTGKPAATTPKEEPTTPKEASVTPKEEPTTPKSAPTTPKEAPTTKSAPTTPKE 196
DB 126 qtdqanryllppppriippcpvcclp-epspp-----pppvtlvt---pqippt 170
OY 197 PAPTTPKEAPTTTPKEAPTTTPKEAPTT---TKSAPTTPKEAPTTTPKKAAPTTPKEAP 254
DB 171 lptlipetpctgppatpctevlpcqptlppqptlppetlppetlppetlppetlppetlpp 226
OY 255 TTPKEEPTTPKEAPTT-KEAPTTKEAPATAKKAAPTTPKEAPTTTPKEAPTTTK 313
DB 227 nlpptpctlppetlppetlppetlppetlppetlppetlppetlppetlppetlppetlpp 286
OY 314 EESPTTPKEAPTTTKSAPTTPKEAPTTTKSAPTTPKEEPTTPKEAPTTTPKEAPTT 373
DB 287 ctsptlpetstpltpetlppetlppetlppetlppetlppetlppetlppetlppetlppetl 340
OY 374 PKKPAPTTPKEAPTTTPKEAPTTTKKKAAPTTPKEAPTTTPKEAPTTTPKEAPTT 433
DB 341 ppgtppntlppetlppetlppetlppetlppetlppetlppetlppetlppetlppetlpp 400
OY 434 APTEPEKPAPTTPKEAPTTPEEPTTPKEAPTTTPKAAAPTTPKEAPTTTPKEAPTT 493
DB 401 lppntppgtpntlppetlppetlppetlppetlppetlppetlppetlppetlppetlpp 460
OY 494 PKPAPTTPKEAPTTTPKKAAPTTPKEAPTTTPKKAAPTTPKEAPTTTPKEAPTT 553
DB 461 pkltp-pltppetlppetlppetlppetlppetlppetlppetlppetlppetlppetlpp 514
OY 554 TPKGTAPTTPKEAPTTTPKEAPTTTPKKAAPTTPKEAPTTTPKKAAPTTPKKAAPTTP 613
DB 515 tppgtppnspqtlppetlppetlppetlppetlppetlppetlppetlppetlppetlpp 574
OY 614 TTSDDP-----APTTPKEAPTTTPKEAPTTTPKKAAPTTPKEAPTTTPKEAPTT 666
DB 575 lppntppgtpntlppetlppetlppetlppetlppetlppetlppetlppetlppetlpp 634
OY 667 P---TTTHKSPDESTPELSAETPKALENSPKKPGVPTTKTAATKP 710
DB 635 pppctllppspstppnspspkspqppppprtfqppppprgpc 681
```

RESULT 9

AAW43106 standard; Protein: 1664 AA.

AAW43106;

16-OCT-1998 (first entry)

C. thermocellum O1pB protein.

Multimer; enzyme; complex; protein-protein interaction; dockerin domain;
cohesin domain; catalytic subunit; scaffold subunit; SdbA; synergistic;
cellulose integrating protein; scaffoldin dockerin binding protein.

Clostridium thermocellum.

Location/Qualifiers

Key 28..192

Domain /note="cohesin type II domain"

Domain /note="cohesin type II domain"

FT

Db 2069 tptptgtgttpttpttlttttvtpttptgtgtpt-tpttlttttvtpttptgtgtpt 2127
QY 669 TIHKSPDESTPELSAEPDPALENSPKEP-----GVPT-TKRPATKPEMTTAK 717
Db 2128 t---tlttttvtpttptgtgtpttpttlttttvtpttptgtgtpttpttltttvt 2184
QY 718 DKTTEDLR--TTP-ETTAAFKMT--KETATTEKTESKITATTOVSTTQDT 769
Db 2185 ptptptgtgttpttlttttvtpttptgtgtpttpttlttttvtpttptgtgtptt 2243
QY 770 TPRKITLTKTTTAPKTYTT-KKITTEIMNKPEETAKKDRATNSKATTPKQKPTKA 828
Db 2244 tp--ltt--lttvtpttptgtgtpttpttlttttvtpttptgtgtpttltttvt 2299
QY 829 P-KKPTSTKRPKTPRVKPKTPTPRKMTSTPELNP--TSRIAMLQTTTR-PNQT 883
Db 2300 ptptptgtgttpttlttttvtpttptgtgtpttpttlttttvtpttptgtgtptt 2359
QY 884 PNSKLVEVNPKSSEDAGAEGETPHMLLRPHVFMPEVTP 921
Db 2360 ptttttvtpttptgt--tqtp--tltpttltttvtpt 2392

RESULT 8
AAG38942
ID AAG38942 standard; Protein; 763 AA.
XX
AC AAG38942;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 48115.
XX
KM Protein identification: signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.
XX
OS Arabidopsis thaliana.
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 07-MAY-1999; 99US-0132486.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145813.
PR 27-JUL-1999; 99US-0145818.
PR 27-JUL-1999; 99US-0145919.

```

Db      157 vknkknrtkkkptkpvvdeagsgldngdfkvttdtctcqnkvstspkittakpin 216
QY      61 PRPSLPNNDTSKETSILYNNKETVTFTKTTNNKOTSDGKEKTSIAKETOSIEETSAK 120
Db      217 prpslpnndtsketsilvnnketvteltkntkqstldgkekttsaketgtsak 276
QY      121 DIAPTSKVLAKPTPKAETTTKGPALTTPKKEPTTTPKKEPASTTPKPPPTTIKASAPTTTK 180
Db      277 diaptskvlakptpkacttktpalttpepnpptpkpcasttkpcttksaptttk 336
QY      181 EPAPTTTSAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTP 240
Db      337 epapttksapttpkpapttkpapttkpapttkpapttkpapttkpapttkpapttkp 396
QY      241 PKKPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKE 298
Db      397 ctcp-----tpkcpnpplprslmptkcpapttkpapttkpapttkpapttkpapttk 443

RESULT 7
AAM24516
ID      AAM24516 standard. Protein: 5179 AA.
XX
AC      AAM24516;
XX
DT      12-OCT-2001 (first entry)
XX
DE      C899P predicted amino acid sequence.
XX
KW      Human; immunotherapy; diagnosis; colon cancer; colon tumour;
XX      immunogenic; gene therapy; vaccine; colonic cancer.
OS      Homo sapiens.
PN      WO200149716-A2.
PD      12-JUL-2001.
PF      29-DEC-2000; 2000WO-US35596.
PR      30-DEC-1999; 9905-0476296.
PR      10-JAN-2000; 2000US-0480321.
PR      15-FEB-2000; 2000US-0504629.
PR      06-MAR-2000; 2000US-0519444.
PR      19-MAY-2000; 2000US-0575251.
PR      29-JUN-2000; 2000US-0609448.
PR      28-AUG-2000; 2000US-0649811.
XX
PA      (CORI-) CORIXA CORP.
PI      Xu J., Lodges MJ, Secretist H, Benson DR, Meagher MJ, Stolk JA;
PI      King GE, Wang T, Jiang Y;
DR      WPI; 2001-441847/47.
XX
PT      Colon tumor associated proteins and nucleic acids useful for the
PT      prevention, diagnosis and treatment of colonic cancer -
XX
PS      Claim 2; Page 446-462; 472p; English.
XX
XX      The present invention describes colon tumour associated proteins (I) and
XX      the polynucleotides (II) that encode them. (I) have cytostatic activity.
XX      (I) and (II) can be used in gene therapy and vaccine production. (I) and
XX      (II) may be used in the prevention, diagnosis and treatment of diseases
XX      associated with inappropriate colon tumour associated protein (TCAP)
XX      expression, such as colonic cancer. For example, (I) and (II) may be
XX      used to treat disorders associated with decreased expression by
XX      rectifying mutations or deletions in a patient's genome that affect the
XX      activity of TCAPs by expressing inactive proteins or to supplement the
XX      patients own production of them. Additionally, (II) may be used to
XX      produce the TCAP proteins, by inserting the nucleic acids into a host
XX      cell culturing the cell to express the protein. (II) and its
XX      complementary sequences may also be used as DNA probes in diagnostic

```

```

CC      polymerase chain reaction (PCR) and hybridisation assays to detect and
CC      quantitate the presence of similar nucleic acids in samples, and
CC      therefore which patients may be in need of restorative therapy. (I) may
CC      also be used as antigens in the production of antibodies against TCAPs
CC      and in assays to identify modulators of TCAP expression and activity.
CC      Anti-(I) antibodies and antagonists may also be used to down regulate
CC      TCAP expression and activity. The anti-(I) antibodies may also be used
CC      as diagnostic agents for detecting the presence of TCAPs in samples
CC      (e.g. by enzyme linked immunosorbant assay (ELISA)). AA128460 to AA129512
CC      and AAM24494 to AAM24523 represent nucleotide and amino acid sequences
CC      given in the exemplification of the present invention.
XX
SQ      Sequence 5179 AA:
XX
Query Match 22.7%; Score 1168; DB 22; Length 5179;
Best Local Similarity 32.5%; Pred. No. 1.8e-57;
Matches 363; Conservative 53; Mismatches 448; Indels 254; Gaps 40;

QY      21 DEAGSGLDNGDFK-----VTPPTSTTQH-NKVSNSPK----- 52
Db      1312 dhpsgsddgdrepgfvgcgapediecrsvkdpnlslsegqgkvqcdvsyglcknedgf 1371
QY      53 -----ITWAKPINRPSLIPNSDTSKETSILYNNKETTYETK 88
Db      1372 gngpfqlydyklrvncwpmckcltspstltpspstltpstltpstltpstltpstltp 1427
QY      89 ETTTNNKOTSDGKEKTSIAKETOSIEKTSKDLAPSKVLAPKAEITGAPALTTP 148
Db      1428 -tttpttpttspstltpstltpstltpstltpstltpstltpstltpstltpstltp 1470
QY      149 KEPT-----PTPKKEPASTTPKKEPTTPTIKASAP-TTPKKEP-----APTTPKSA 194
Db      1471 sppttspstltpstltpstltpstltpstltpstltpstltpstltpstltpstltpstl 1530
QY      195 KEKPAPTTPKKEPAPTTPKKEP-----APTTPKKEPAPTTPKSA--TTPKKEPAPTTPK 245
Db      1531 ttppttpttspstltpstltpstltpstltpstltpstltpstltpstltpstltpstl 1590
QY      246 PTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEP-----APTAKKAPPTTPKE 298
Db      1591 itttpttpttspstltpstltpstltpstltpstltpstltpstltpstltpstltpstl 1650
QY      299 PAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKSAPTTPKKEPAPTTPKSA--TTPKKEP 355
Db      1651 pppttspstltpstltpstltpstltpstltpstltpstltpstltpstltpstltpstl 1709
QY      356 TTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTP 414
Db      1710 ttttpttpttspstltpstltpstltpstltpstltpstltpstltpstltpstltpstl 1769
QY      415 -----KETAPTTP----- 427
Db      1770 fspstltpstltpstltpstltpstltpstltpstltpstltpstltpstltpstltp 1829
QY      428 -----TTPKELAPT 437
Db      1830 dvplgdlqgtvcdvsyglcknedqkpgyvmalfcfnlyinvqcecvctqptmtttt 1889
QY      438 PEKPAPTTPPEELAPTTPPEEPPT-TPEEPAPTTPKAAAPTTPKKEPAPTTPKEP-----AP 491
Db      1890 temppttpttltttlvtpcpqgqtpcttltttlvtpcpqgqtpcttltttlvtpcpqgqtp 1949
QY      492 TTPKKEPAPT-TPEKTAATTPKGTAPTLKAPAPPTTPKKAPELAPTTPPESTSDKP 550
Db      1950 tttvtpcpqgqtpcttltttlvtpcpqgqtpcttltttlvtpcpqgqtpcttltttlvtpcp 2009
QY      551 AP-TTPKGTAPTPKKEPAPTTPKKEPAPTTPKGTAPTLKAPAPPTTPKKAPELAPTTPK 609
Db      2010 tpttpttltttlvtpcpqgqtpcttltttlvtpcpqgqtpcttltttlvtpcpqgqtp 2068
QY      610 GPTSTSDKAP-TTPKETAAPTTPKKEPAPTTPKKAPELAPTTPPEPTTPPTTSEVSTPTTKEPT 668

```

CC particular, one of the surfaces is pericardial tissue. DNA encoding a
 CC tribonectin may be used in gene therapy. The present sequence represents
 CC a substantial portion of a human MSF-derived tribonectin.

XX Sequence 902 AA:

Query Match 57.4%; Score 2956.5; DB 22; Length 902;
 Best Local Similarity 72.2%; Pred. No. 9,9e-159;
 Matches 699; Conservative 28; Mismatches 80; Indels 161; Gaps 65;

QY 1 VKDKKKRRTKKKPPKPPVVDGSGLDNGDFKVTTPDSTTOHKNVSTSPKTTAKPIN 60
 DB 1 vkdkknrrtkkppkppvvdgsgldngdfkvttpdsttqhknvstspkttakpin 60
 QY 61 PRPSLPNSDTSKETS/LTVNKKETVETKETTNTKOTSGGKEKTKAKENOSTEKTSAK 120
 DB 61 prpslpnsdtsketsltvnkketvettntkotsggkeltkakevsgtsktsak 120
 QY 121 DLAPTSTVLAKEPTPKAETTTGPA/LTTPKKEPTTPPKKPAETTPKKEPTTPKSA 180
 DB 121 dlaptstvlakepstkpaetttgpa/lttppkkpattppkkpattppkksa 180
 QY 181 EPAPTTKSAPTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPKKEAPTT 240
 DB 173 epapttksapttppkkeapttppkkeapttppkkeapttppkkeapttppkkeap 240
 QY 241 PKRPAPTPKKEAPTTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPKKEA 300
 DB 229 -keappt-keappt-keappt-keappt-keappt-keappt-keappt-keappt-keap 300
 QY 301 PTPPKKEAPTTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPKKEAPTT 360
 DB 282 ptt-keap-tpk-keappt-keappt-keappt-keappt-keappt-keappt-keap 360
 QY 361 PAPTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPKKEAPT 420
 DB 336 paptt-keappt-keappt-keappt-keappt-keappt-keappt-keappt-keappt 420
 QY 421 TPKKLPTTPPEKLAPTTPPEKAPTTPEELAPTTPEEPTTPPEKAPTTPEEPTTPPE 480
 DB 389 t-keappt-keappt-keappt-keappt-keappt-keappt-keappt-keappt-ke 480
 QY 481 PAPTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPKKEAPT 540
 DB 441 paptt-keappt-keappt-keappt-keappt-keappt-keappt-keappt-keap 540
 QY 541 EPTSTSDKAPTTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPKKEAPT 600
 DB 489 epapttk-keappt-keappt-keappt-keappt-keappt-keappt-keappt-keap 600
 QY 601 KELAPTTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPKKEA 660
 DB 542 -----tkkeapptke-keappt-keappt-keappt-keappt-keapptke-keap 660
 QY 661 PTTTKSEPTTHKSPDSTPELSAEPTRKALENSPKRGVTTTPKPAATKEMTTTAKDKT 720
 DB 590 -pttksepttkkeappt-keappt-keappt-keappt-keappt-keapptke-keap 720
 QY 721 TEHDLRTPEPTTAARKMETATTEKTESKTTATTTTODTTPKTTTTLKT 780
 DB 637 -----papt-----tkkeappt-keappt-keapptkeap----- 780
 QY 781 TLAPKVTTKTTTTEINMKPEETAKPKDRATNSKATTPKPKPKAPKSTTKPKPT 840
 DB 668 -----tlk-keapptke-keappt-keappt-keappt-keappt-keappt-keap 840
 QY 841 MPKVRKPTTPTRPKMTSTWELNPTSRLEAMLOTTRNPNPNKIVNVNKSSEDAGS 900
 DB 687 -----aptpckmtstwelnptsrleamlottrnppnpnkivnvnskssedag 900
 QY 901 AEGETPHMLRPVFMPEVTPMDVLPVRVNOGIIINPMISDETNCNGKPVGLTTLRN 960
 DB 901 aegetphmlrpvfmpevtpmdvlpvrvnoგიიინპმისდეტნცნგკპვგლტტლრნ 960

DB 740 aegtpthmlrphvfmpvtpmdvlpvrpqgllnplmsdetnclngkpvglttlarn 799
 QY 961 GTLVAFRG 968
 DB 800 gtlvafrg 807

RESULT 6
 AAR80041
 ID AAR80041 standard; Protein; 452 AA.
 XX
 AC AAR80041;
 XX
 DT 10-Apr-1996 (first entry)
 XX
 DE Human megakaryocytopoietin protein.
 XX
 KW Human; megakaryocytopoietin; wheat germ agglutinin; heparin;
 KM megakaryocyte; aplastic anaemia; thrombocytopenia; hematocytopenia;
 KM multipotential stem cell.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 393..396
 FT /note="unspecified amino acids"
 FT Misc-difference 444..446
 FT /note="unspecified amino acids"
 XX
 PN W09523861-A1.
 XX
 PD 08-SEP-1995.
 XX
 PE 06-MAR-1995; 95MO-CN00015.
 XX
 PR 04-MAR-1994; 94CN-0112066.
 XX
 PA (SHAN-) SHANGHAI BEITI BIOTECHNOLOGY CO LTD.
 XX
 PI Gu X, Han Z, Shen Q;
 XX
 DR MPI, 1995-320576/41.
 DR N-PSDB; AAT04546.
 XX
 PT New haematopoietic cell growth factor - used for treating
 XX thrombocytopenia and hematocytopenia
 PS Example: Page 23; 36pp; Chinese.
 CC This sequence represents the human megakaryocytopoietin (MPO) protein.
 CC This sequence was purified using a carrier which can couple wheat germ
 CC agglutinin and heparin to separate MPO. Fragments of this sequence (see
 CC AAR80039 and AAR80040) were used to produce the amplification primers
 CC shown in AAT04544 and AAT04545. The fragments amplified by these primers
 CC can then be used as probes to screen human cDNA libraries for MPO cDNA.
 CC The MPO cDNA can then be inserted into a plasmid which is used to
 CC transform cells to produce MPO. The MPO sequence is capable of promoting
 CC colony formation of megakaryocytes, enlarging the size of megakaryocytes
 CC and stimulating the proliferation of multipotential stem cells. The
 CC factor may be used for treating thrombocytopenia and hematocytopenia.
 CC The purification method can be used to isolate MPO from human urine or
 CC serums of patients with aplastic anaemia, and from animal blood or urine
 CC by radiation exposing the animals to induce aplastic anaemia.
 XX
 SQ Sequence 452 AA:

Query Match 26.3%; Score 1357.5; DB 16; Length 452;
 Best Local Similarity 88.3%; Pred. No. 3.7e-69;
 Matches 263; Conservative 3; Mismatches 21; Indels 11; Gaps 1;
 QY 1 VKDKKKRRTKKKPPKPPVVDGSGLDNGDFKVTTPDSTTOHKNVSTSPKTTAKPIN 60
 DB 1 vkdkknrrtkkppkppvvdgsgldngdfkvttpdsttqhknvstspkttakpin 60

CC measuring the amount of MSF or its fragment in a biological sample of a
 CC mammal, wherein an increased amount of MSF compared to a control
 CC indicates the presence of or predisposition to developing
 CC osteoarthritis. The tribonectin and DNA encoding it are useful in the
 CC treatment of osteoarthritis, where they may be used for lubricating
 CC mammalian joints, such as articulating joints of humans, dogs or horses.
 CC The tribonectin, when formulated as a membrane, foam, gel or fibre, is
 CC useful for inhibiting adhesion between two surfaces such as the injured
 CC tissues of a mammal, where the injury is caused by a surgical insertion
 CC or trauma, or an artificial device e.g., an orthopaedic implant. In
 CC particular, one of the surfaces is pericardial tissue. DNA encoding a
 CC tribonectin may be used in gene therapy. The present sequence represents
 CC human MSF.
 CC
 XX
 XQ Sequence 1404 AA;

Query Match	100.0%	Score 5155;	DB 22;	Length 1404;
Best Local Similarity	100.0%;	Pred. No. 4.2e282;		
Matches 968;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

Oy	1	VKDNNKNTKPKPPKPPVVDGASGJLNDGDEVTTIPDSTTQHNKVSTSPKTTAKPIN	60
Db	200	Vkdnkknrtkkkpcpdpvndaagsqldngdfkvrttcdstltbhnkvsspxklttkapn	259
Oy	61	PRPSLPRNSDSKESLSLVNKEVTEVKEKTTTNNKOPSTGKSKTTSAKETOSIEKTSAK	120
Db	260	prpslprnsdstsketsltvnkettvetkeltcnkqtsldgktsaketsaketsaketsak	319
Oy	121	DLAPTSKVLAHPKPAETTTKGPALTTPKBEPPTPKBEPPTPKBEPSTTPKBPPTTIKSAPTPK	180
Db	320	dlaptskvylakpckaeittkqpaltttkpceppttkpcepsttkpceppttkksaptpk	379
Oy	181	EPAPITTKSAPTTKKEAPATTTTKKBPATTTKKEAPATTTTKKBPATTTTKKSAPTTKKEAPATTT	240
Db	380	epapittksapttkkepapttkkbpatttkkbpatttkkbpatttkksaptpkcapitk	439
Oy	241	PKKPAPTPPKBPAPTPPKBPPTTPPKBPAPTPPKBPATTTKBPAPTAPKKBPATTTKBP	300
Db	440	pkkbpaptpkbpaptpkbppttpkbpaptpkbpatttkkbpatttkkbpaptpkbpaptpkbp	499
Oy	301	PTTPKEAPATTTKBPSTTPKKEAPATTTTKKSAPTTKKEAPATTTKSAPTTPKBPSPTTYKE	360
Db	500	pttpkepapttkbpspttpkkepapttkksapttkkepapttkksaptpkbpsttpk	559
Oy	361	PAPTPPKBPAPTTPKKBPATTPPKBPAPTTTKKBPATTTTKKBPATTAHPKBPATTPK	420
Db	560	paptpkbpaptpkbpaptpkbpaptpkbpaptpkbpaptpkbpaptpkbpaptpkbpap	619
Oy	421	TPKKTTPTPPKKLAHTTPPEKBPATTPPEBLAHTTPPEEBPTTPTPPEBPATTPPKAAPTTPKE	480
Db	620	tpkkttptpkklahthtppekbpattppelbaththppeebpttpthppebpattpkapaaptpk	679
Oy	481	PAPTPPKBPAPTTKKEAPATTPKKTAPTPPKGTAHTTLKBPAPTTPKKAPKELAPTTTK	540
Db	680	paptpkbpaptpkkepattpkktatpplkgtahthtlkbpaptpkpkapkelaptpk	739
Oy	541	EPTSTTSKBPAPTPPKGAPATTPPKBPAPTPPKBPAPTTPKGTAHTTLKBPAPTTPKKBPAP	600
Db	740	epsttskbpaptpkgtaptpkbpaptpkbpaptpkbpaptpkgtahthtlkbpaptpkbpap	799
Oy	601	KELAPTTTKGATSTTSDBKBPATTPKBPATTPPKBPAPTPPKBPAPTPPKBPATTPPKBPATTP	660
Db	800	kelaptpkgtsttsdbkbpattpkbpattpkbpaptpkbpaptpkbpaptpkbpaptpkbpap	859
Oy	661	PTTTPKEPTTIHKSPEDESTPELSABPTTPKPALENSPKBEGVPTTKTPAATKPEMTTTAKDKT	720
Db	860	ptttpkepttihkspdestpelsabpttpkpaalenspkbeqvpttktpaatkpemtttakdk	919
Oy	721	TERDRTTPPETTTAPKPKTKETATTTKETSKTITATTTQVTSSTTMODTTPKKTITLTKTT	780
Db	920	terdrttppetttaapmktketattketskittlatitqvtssttmodttpkktitlttk	979

OY	781	TLAKRVTTTKKTTTITTTTINMKPEETAKPKDRAINSAATTPKPOKPKPAKPKTSRKAKT	840
Db	980	tlakrvtttkkllltlftlmkpeetakpkdranskatlpkpkpkapkpustckpct	1035
OY	841	MPVRKPKTPTTPPKMNSTJMPBLNPTSRIAEAMIQTTTRPNQTPNSKLVENVPKSEDAG	900
Db	1040	mpvrpkktpttpprkmsstmpelnpncrsilaeanlqtltrpnqtpnksklevnpkseag	1095
OY	901	AEGETPHMLRPHVEMPEVTPDMDYLPKVPNOGIIITINPMISDETINCSNKPVDGILTLRN	960
Db	1100	aegetphmlrlphvfmpevpdmndylprvpngiilnmpisdetnlnsgkpvdgiltlrn	1155
OY	961	GTIVAFRG 968	
Db	1160	gtivafrg 1167	

RESULT	5
AAB29778	
ID	AAB29778 standard; Protein; 902 AA

AC AAB29778;

DT 28-FEB-2001 (first entry)

Human MSF-derived tribonectin.

Human tribonectin; MSF; megakaryocyte stimulating factor; KW alternative splicing; joint boundary lubricant; O-linked oligosaccharide

KW osteoarthritis; tribosupplementation; tissue adhesion inhibition;
KW friction coefficient reduction; gene therapy; antiarthritic;
KW osteopathic.

OS Homo sapiens.

PN W0200064930-A2.

PD 02-NOV-2000.

PF 24-APR-2000; 2000WO-US10953.

PR 23-APR-1999; 99US-0298970.

PA (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER

PI Jay GD;

DR WPI; 2001-024673/03.

PT Novel tribonectin polypeptide useful as lubricant for treating

PS Disclosure; Fig 1; 47pp; English.

CC The invention relates to a human tribonectin which is a product of

gene. The tribonectin has at least one O-linked oligosaccharide

CC repeats of a motif having at least 50% identity to the sequence KEPAPTT

CC human MSF-derived tribonectin; a biocompatible composition comprising a

CC of diagnosing osteoarthritis or a predisposition to osteoarthritis by

CC mammal, wherein an increased amount of MSF compared to a control

osteoarthritis. The tribonectin and DNA encoding it are useful in the

mammalian joints, such as articulating joints of humans, dogs or horses.

useful for inhibiting adhesion between two surfaces such as the injured

or trauma, or an artificial device e.g., an orthopaedic implant. In

XX Example 1: Page -: 34pp: English.
 PS
 CC Administration of a composition comprising the campodactylly-arthropathy-
 CC coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.
 CC The composition may further comprise a local anaesthetic. The composition
 CC of the invention may be administered via intra-articular or intravenous
 CC injection. The human CACP protein is identified in the invention as
 CC being megakaryocyte stimulating factor (MSF). The gene encoding
 CC CACP protein (MSF) is located on chromosome 1q25-31, and mutations in
 CC this gene are responsible for the heritable disorder campodactylly-
 CC arthropathy-coxa vara-pericarditis, in which patients have synovial
 CC hyperplasia without evidence of inflammation. CACP protein (MSF)
 CC acts as a synovium lubricant, and can be used to lubricate tissue and
 CC joints in the treatment of osteoarthritis. The composition may be
 CC applied to reduce the symptoms of osteoarthritis (e.g., joint pain,
 CC loss of range of movement or joint damage). The present sequence
 CC represents human megakaryocyte stimulating factor (MSF, CACP protein).
 CC Note: This sequence is not given in its entirety in figure 4 of the
 CC specification, although a GenBank accession number was given. This
 CC sequence was therefore obtained from GenBank (U70316).
 CC
 XX
 SQ Sequence 1404 AA:

Query Match 100.0%; Score 5155; DB 22; Length 1404;
 Best Local Similarity 100.0%; Pred. No. 4.2e-282;
 Matches 968; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKDKKKRKKKKPPKPPVYDAGSGLDNGDFKVTTPDSTGTHNKVSTSPKITTAKPIN 60
 DB 200 VKDKKKRKKKKPPKPPVYDAGSGLDNGDFKVTTPDSTGTHNKVSTSPKITTAKPIN 259
 QY 61 PRPSLPNSDTSKSTLVNKKETVETKETTNTKOSTGKKETSAKKEOSTEKRSK 120
 DB 260 PRPSLPNSDTSKSTLVNKKETVETKETTNTKOSTGKKETSAKKEOSTEKRSK 319
 QY 121 DLAPTSTVLAKPPPKAETTTGKALTTPKKEPTTPPKKPASTTPKKEPTTPPK 180
 DB 320 DLAPTSTVLAKPPPKAETTTGKALTTPKKEPTTPPKKPASTTPKKEPTTPPK 379
 QY 181 EPAPTTKSAPTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPK 240
 DB 380 EPAPTTKSAPTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPK 439
 QY 241 PKRPAPTTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPK 300
 DB 440 PKRPAPTTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPK 499
 QY 301 PTPPKKEAPTTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPK 360
 DB 500 PTPPKKEAPTTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPK 559
 QY 361 PAPTPKKEAPTTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPK 420
 DB 560 PAPTPKKEAPTTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPK 619
 QY 421 TPKKLTPTTEKLAPTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPK 480
 DB 620 TPKKLTPTTEKLAPTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPK 679
 QY 481 PAPTPKKEAPTTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPK 540
 DB 680 PAPTPKKEAPTTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPKKEAPTTTPK 739
 QY 541 EPSTSTSDKRAPTPPKGATATTPKKEAPTTTPKKEAPTTTPKKEAPTTTPK 600
 DB 740 EPSTSTSDKRAPTPPKGATATTPKKEAPTTTPKKEAPTTTPKKEAPTTTPK 799
 QY 601 KELAPTTTPKSTSDKRAPTPPKGATATTPKKEAPTTTPKKEAPTTTPKKEAPTTTPK 660
 DB 800 KELAPTTTPKSTSDKRAPTPPKGATATTPKKEAPTTTPKKEAPTTTPKKEAPTTTPK 859

QY 661 PTTTKKEPTTIHKSPDESTPELSAEPTRKALENSRKEPQVPTTKTPAATKPEMTTANDKT 720
 DB 860 PTTTKKEPTTIHKSPDESTPELSAEPTRKALENSRKEPQVPTTKTPAATKPEMTTANDKT 919
 QY 721 TERLRTPTTEPTTAPKMTKETATTPKKEPTTIKATTTTQVSTTQDDTPPEKITTAKTT 780
 DB 920 TERLRTPTTEPTTAPKMTKETATTPKKEPTTIKATTTTQVSTTQDDTPPEKITTAKTT 979
 QY 781 TLAPKVTTKTKTTTTEIMNKKPEETANPKDRAATNSKATTPKPKPKPKSTTKPKPT 840
 DB 980 TLAPKVTTKTKTTTTEIMNKKPEETANPKDRAATNSKATTPKPKPKPKSTTKPKPT 1039
 QY 841 MPVRKKTPTTPPKKMTSTWELNPTSRISAEAMLTTPRPNQTNLSKLVENPKSEDAAG 900
 DB 1040 MPVRKKTPTTPPKKMTSTWELNPTSRISAEAMLTTPRPNQTNLSKLVENPKSEDAAG 1099
 QY 901 AEGETPHMLRPHVFNPEVTPDMQDLPRVPNQGIIIPMLSDFTNIGNKRVPDGLTTLRN 960
 DB 1100 AEGETPHMLRPHVFNPEVTPDMQDLPRVPNQGIIIPMLSDFTNIGNKRVPDGLTTLRN 1159
 QY 961 GTLVAFNG 968
 DB 1160 GTLVAFNG 1167

RESULT 4
 AAB29773
 ID AAB29773 standard; Protein: 1404 AA.
 XX
 AC AAB29773;
 XX
 DT 28-FEB-2001 (first entry)
 XX
 DE Human megakaryocyte stimulating factor (MSF), SEQ ID NO:1.
 XX
 KW Human MSF: megakaryocyte stimulating factor; tribonectin;
 KW alternative splicing; joint boundary lubricant; O-linked oligosaccharide;
 KW osteoarthritis; tribosupplementation; tissue adhesion inhibition;
 KW friction coefficient reduction; gene therapy; antiarthritic;
 KW osteopathic.
 XX
 OS Homo sapiens.
 XX
 PN MO200064930-A2.
 XX
 PD 02-NOV-2000.
 XX
 PF 24-APR-2000; 2000MO-US10953.
 XX
 PR 23-APR-1999; 99US-0298970.
 XX
 PA (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.
 XX
 PI Jay GD;
 XX
 DR WPI: 2001-024673/03.
 XX
 DR N-FSDB: AAC81498.
 XX
 PT Novel tribonectin polypeptide useful as lubricant for treating
 XX osteoarthritis, comprises O-linked lubricating moiety
 XX
 PS Claim 3; Page 7; 47pp: English.
 CC
 CC The invention relates to a human tribonectin which is a product of
 CC alternative splicing of the human MSF (megakaryocyte stimulating factor)
 CC gene. The tribonectin has at least one O-linked oligosaccharide
 CC lubricating moiety and has a polypeptide sequence comprising 1-76
 CC repeats of a motif having at least 50% identity to the sequence KEKAPPT
 CC (AAB29774). The invention also relates to a nucleic acid encoding a
 CC human MSF-derived tribonectin; a biocompatible composition comprising a
 CC human tribonectin for inhibiting tissue adhesion formation; and a method
 CC of diagnosing osteoarthritis or a predisposition to osteoarthritis by

DR WPI: 1992-284660/34.
DR N-PSDB: AA027223.

XX New human mega-karyocyte stimulating factors - for treating
PT immune deficiencies, cancer, exposure to radiation or drugs,
PT bacterial and viral infections, etc.

PS Claim 1, 2 and 3; Fig 1; 87pp; English.

XX The sequence given is a full length translation from the megakaryocyte
CC stimulating factor (MSF) precursor. The sequence covered by exons II,
CC III and IV encodes megakaryocyte stimulating factor (MSF). This
CC sequence is modified by the addition of an N-terminal sequence encoding
CC a secretory leader, an initiating methionine preceding exon II and a
CC terminating codon following exon IV. The cDNA sequence given contains
CC sequences derived from human megakaryocyte colony stimulating factor
CC (meg-CSF). Exon I contains the initiating methionine, and encodes a
CC classical mammalian protein secretion signal sequence. The sequence
CC encoding the original meg-CSF includes exons II-IV and is thought to
CC terminate in the region between amino acid residues 134 - 147. The
CC primary transcript of this gene may be cleaved in different ways to
CC yield a family of mRNA's each encoding a different MSF protein. Exons
CC V and VI are thought to be related to the activity of the factor and
CC are also implicated in the stability, folding and processing of the
CC molecule. These exons are also thought to play a role in the observed
CC synergy of MSF with other cytokines. Exons V - XII are believed to be
CC implicated in the processing or folding of the appropriate structure of
CC the resulting factor, i.e. one or more of these exons may contain
CC sequences which direct proteolytic cleavage, adhesion, organisation of
CC the cellular matrix or extracellular matrix processing. Both naturally
CC occurring and non-naturally occurring MSF's may be characterised by
CC various combinations of alternatively spliced exons from this sequence,
CC with the exons spliced together in differing orders to form different
CC members of the MSF family.

CC Sequence 1404 AA:

Query Match 100.0%; Score 5155; DB 13; Length 1404;
Best Local Similarity 100.0%; Fred. No. 4.2e-282;
Matches 968; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKDNKKNRRTKKRPKPPVVDGSGLDNGDFVTTPTDSTOHNNKYSTPKITTTAKPIN 60
DB 200 vkdkknrrtkkrrppvvdsgslndgdfvtptdstohnnkxstpskltakpin 259
QY 61 PRPSLPPNSDTSKETSILVNNKETTVETKETTNNKQSTIDGKERTTSAKTOSTIEKTSAK 120
DB 260 prpslppnsdtsketsiltvnnketvtetkettlnkqstidgkertsaketsaksak 319
QY 121 DLAPTGSVLAKPPPKAETTKGPAITTPKEPTTPKEPASTPKKEPTPTTISAPPTPK 180
DB 320 dlaptgsvlakpppkatettkgpaittpkepttpkepastpkkeptpttiksapptpk 379
QY 181 EPAPTTTSAPPTPKKEPAPTTTKEPAPTTTKEPAPTTTTSAPTTTKEPAPTT 240
DB 380 epaptttsapptpkkepapttttkepapttttkepapttttsappttkepaptt 439
QY 241 PKRPAPTTKEPAPTTKEPTPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 300
DB 440 pkrappttkepapttkeptpttpkepapttttkepapttttkepapttkepapa 499
QY 301 PPTPKKEPAPTTKEPSTPTTPKEPAPTTTTSAPTTTKEPAPTTTTSAPTTKEP 360
DB 500 ptpkkepapttkepstpttpkepapttttsappttkepapttttsappttke 559
QY 361 PAPPTKEPAPTTPKKAPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 420
DB 560 papptkepapttpkkapapttttkepapttttkepapttttkepapttkepapt 619
QY 421 TPKKLTPTTEKLAAPTPEKPAAPTTPEELAPTTPEEPTPTTPEEAPATTTKAANPTKE 480
DB 620 tpkkltptteklatptpekpaapttpeelapttpeeptpttpeeapatttkaanptke 679

QY 481 PAPPTPKKEPAPTTPKKAPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 540
DB 680 papptpkkepapttpkkapapttttkepapttttkepapttttkepapttkepapt 739
QY 541 EPTSTSDKPAAPTPKGAATTPPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 600
DB 740 eptstsdkapapttpkgaattppkkepapttttkepapttttkepapttkepapt 799
QY 601 KELAPTTTSGPSTTSDKPAAPTTPKETAPTTTKEPAPTTTKEPAPTTTKEPAPTT 660
DB 800 kelaptttsgpsttsdkpapttpketapttttkepapttttkepapttkepapt 859
QY 661 PTTTKEPTTTHKSPDSTPELSAEPPTKALENSKPEGVPTTPATKPEMTTAKDKPT 720
DB 860 ptttkepttthkspdstpeelsaepptkalenskpegvpttpatkpemttakdkpt 919
QY 721 TERDLRTPEETTAARKMTKETATTEKTESKITATTTQVSTTQDTTPFKITTLKTT 780
DB 920 terdlrtpeettaarkmktetattekteskiteskitatttqgstttdtptfkittlkt 979
QY 781 TLAPKVTYTKTITTTTBEIMNKKPEETAKPKDRATNSKATTPKPKPKARKKPTSTKKPT 840
DB 980 tlapkvttytktittttbeimnkkpeetakpkdratnskattpkpkpkarkkptstkkpt 1039
QY 841 MPVRKPKTTPPRKMTSTWPELNPSTRIAEAMLOTTTPRPNQNSKIVEVNPSEDAAG 900
DB 1040 mpvrkpkttpprkmtstwpeelnpstriaeamlottpprpnqnsklivevnpksedag 1099
QY 901 AEGETPHMLRPVMEVTPDMDYLPVPNGGIIINPMLSDETINCNCKPVDGTLTLRN 960
DB 1100 aegetphmlrpvmevtpdmfylpvpnggiinpmldsdetinncnckpvdgtltrlrn 1159
QY 961 GTLVAFRG 968
DB 1160 gtlvafrg 1167

RESULT 3

AAB60568
ID AAB60568 standard; protein; 1404 AA.

AC AAB60568;
XX

DT 27-APR-2001 (first entry)

XX Human megakaryocyte stimulating factor (MSF, CACP).

KW Human; CACP protein; camplodactylly-arthropathy-coxa vara-pericarditis;

KW MSF; megakaryocyte stimulating factor; synovial lubricant;

KW chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;

XX antiarthritic.

OS Homo sapiens.

PN WO200107068-A1.

XX 01-FEB-2001.

PF 21-JUL-2000; 2000WO-US20002.

XX 19-JUL-1999; 9905-0145328.

PR 13-JUL-2000; 2000US-0145328.

PA (UYCA-) UNIV CASE WESTERN RESERVE.

PI Warman MI;

PT New composition comprising the camplodactylly-arthropathy-coxa
PT vara-pericarditis protein in combination with an anesthetic, useful for
PT treating osteoarthritis, or as lubricants of tissue and joints -

DR N-PSDB: AAH98981.
 XX Isolated polypeptide for treatment of diseases, diagnostics, raising
 PT antibodies and research use -
 PT
 XX
 PS Claim 20: Page 1198-1201; 1275pp; English.
 XX
 CC The present invention provides the protein and coding sequences of novel
 CC proteins from a variety of organisms, including human, dog, cat, horse,
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
 CC from the organism of interest. They can be used in diagnostics,
 CC forensics, gene mapping, identification of mutations, to assess
 CC biodiversity and for nutritional purposes. The present sequence is a
 CC protein of the invention.
 CC
 XX
 SO Sequence 1299 AA:

Query Match 100.0%; Score 5155; DB 22; Length 1299;
 Best Local Similarity 100.0%; Pred. No. 3,9e-282;
 Matches 968; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKDNKKNTKKKPPKPVVDAGSGLDNGDFKVTTPDTSTTQHKKVSTSKITAKPIN 60
 DB 200 VKDNKKNTKKKPPKPVVDAGSGLDNGDFKVTTPDTSTTQHKKVSTSKITAKPIN 259
 QY 61 PRPSLPNSDTSKESLTVNKETVETKETTNNKOTSDGKEKTSKESQSTESAK 120
 DB 260 PRPSLPNSDTSKESLTVNKETVETKETTNNKOTSDGKEKTSKESQSTESAK 319
 QY 121 DLATSKVLAKPTPKAETTTGPAALTPPKETPTTPKEAPASTPKETPTTKSAPTPK 180
 DB 320 DLATSKVLAKPTPKAETTTGPAALTPPKETPTTPKEAPASTPKETPTTKSAPTPK 379
 QY 181 EPAPTTKSAATPKKEAPATTKEAPATTKEAPATTKEAPATTKEAPATTKEAPATT 240
 DB 380 EPAPTTKSAATPKKEAPATTKEAPATTKEAPATTKEAPATTKEAPATTKEAPATT 439
 QY 241 PKKPAPTPKKEAPATTKEAPATTKEAPATTKEAPATTKEAPATTKEAPATTKEA 300
 DB 440 PKKPAPTPKKEAPATTKEAPATTKEAPATTKEAPATTKEAPATTKEAPATTKEA 499
 QY 301 PTPKKEAPATTKEAPATTKEAPATTKEAPATTKEAPATTKEAPATTKEAPATTKE 360
 DB 500 PTPKKEAPATTKEAPATTKEAPATTKEAPATTKEAPATTKEAPATTKEAPATTKE 559
 QY 361 PAPTPKKEAPATTKEAPATTKEAPATTKEAPATTKEAPATTKEAPATTKEAPATT 420
 DB 560 PAPTPKKEAPATTKEAPATTKEAPATTKEAPATTKEAPATTKEAPATTKEAPATT 619
 QY 421 TPKKLPTTTEKLAAPTTPEKPAAPTTPEELAPTTPEEPPTTPPEEPAPTPKAAAPNTPE 480
 DB 620 TPKKLPTTTEKLAAPTTPEKPAAPTTPEELAPTTPEEPPTTPPEEPAPTPKAAAPNTPE 679
 QY 481 PAPTPKKEAPATTKEAPATTKEAPATTKEAPATTKEAPATTKEAPATTKEAPATT 540
 DB 680 PAPTPKKEAPATTKEAPATTKEAPATTKEAPATTKEAPATTKEAPATTKEAPATT 739
 QY 541 EPSTSTSDKAPPTPKGTAPTPKEAPATTKEAPATTKEAPATTKEAPATTKEAPATT 600
 DB 740 EPSTSTSDKAPPTPKGTAPTPKEAPATTKEAPATTKEAPATTKEAPATTKEAPATT 799
 QY 601 KELAPTTTGGPTSTTSKAPATTPEKPAAPTTPEELAPTTPEEPPTTPPEEPPTTS 660
 DB 800 KELAPTTTGGPTSTTSKAPATTPEKPAAPTTPEELAPTTPEEPPTTPPEEPPTTS 859
 QY 661 PTTTKETPTTHKSPDSTPELSAEPTPKALENSPKEGVTTKTPAATKEMTTTKDKDT 720
 DB 860 PTTTKETPTTHKSPDSTPELSAEPTPKALENSPKEGVTTKTPAATKEMTTTKDKDT 919
 QY 721 TENDLRTPTPTTTAAKMKETATTTETTESKITATTTQVSTTQDTTPFKITLTKTT 780

DB 920 terdlrttpeetlaapkmktelatttektlesklatattqvtsttqdtcpfkltlclt 979
 QY 781 TLAPKVVTTKKKVVTTTTTEINMKPEETAKPKDRATNSKATTPKPKPKKPKTKPKKT 840
 DB 980 TLAPKVVTTKKKVVTTTTTEINMKPEETAKPKDRATNSKATTPKPKPKKPKTKPKKT 1039
 QY 841 MPVRKPKTTTPPKKMTSTPELNFTSKRIADAMLOTTTRPNQTPNSKLVEVNPKSEDAG 900
 DB 1040 MPVRKPKTTTPPKKMTSTPELNFTSKRIADAMLOTTTRPNQTPNSKLVEVNPKSEDAG 1099
 QY 901 AEGEPHMLLRPHVEMPEVTPDMDLPRVPMOGIIINPMISDENINCKRVODGLTTRN 960
 DB 1100 AEGEPHMLLRPHVEMPEVTPDMDLPRVPMOGIIINPMISDENINCKRVODGLTTRN 1159
 QY 961 GTLVAFRG 968
 DB 1160 gtlvafrg 1167

RESULT 2
 AAR26049
 ID AAR26049 standard; Protein: 1404 AA.
 XX
 AC AAR26049;
 XX
 DT 02-FEB-1993 (first entry)
 XX
 DE MSF precursor.
 XX
 KW Megakaryocyte colony stimulating factor; secretion signal; meg-CSF;
 KM stability; proteolytic cleavage; adhesion; alternative splicing.
 XX
 OS Synthetic.
 XX
 FH Key
 FT Region
 FT /label= Exon_I
 FT 26..67
 FT /label= Exon_II
 FT 67..107
 FT /label= Exon_III
 FT 107..157
 FT /label= Exon_IV
 FT 157..200
 FT /label= Exon_V
 FT 200..1141
 FT /label= Exon_VI
 FT 1411..1166
 FT /label= Exon_VII
 FT 1166..1212
 FT /label= Exon_VIII
 FT 1213..1266
 FT /label= Exon_IX
 FT 1266..1331
 FT /label= Exon_X
 FT 1331..1373
 FT /label= Exon_XI
 FT 1373..1404
 FT /label= Exon_XII
 FT
 PN WO9213075-A.
 PD
 PD 06-AUG-1992.
 XX
 PF 17-JAN-1992; 92WO-US00433.
 XX
 PR 18-JAN-1991; 91US-0643502.
 PR 10-SEP-1991; 91US-0757022.
 XX
 PA (GENW) GENETICS INST INC.
 XX
 PI Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K;
 XX

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:16:03 ; Search time 107.17 Seconds
(without alignments)
669.058 million cell updates/sec

Title: US-09-556-246-1_COPY_200_1167

Perfect score: 5155
Sequence: 1 VKDKKKRKKKKPPKPPV.....GKPVDTLTTLNGLTVAFRG 968

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101.*

1:	/SIDSR/gcgdata/geneseq/geneseq/AA1980.DAT.*
2:	/SIDSR/gcgdata/geneseq/geneseq/AA1981.DAT.*
3:	/SIDSR/gcgdata/geneseq/geneseq/AA1982.DAT.*
4:	/SIDSR/gcgdata/geneseq/geneseq/AA1983.DAT.*
5:	/SIDSR/gcgdata/geneseq/geneseq/AA1984.DAT.*
6:	/SIDSR/gcgdata/geneseq/geneseq/AA1985.DAT.*
7:	/SIDSR/gcgdata/geneseq/geneseq/AA1987.DAT.*
8:	/SIDSR/gcgdata/geneseq/geneseq/AA1987.DAT.*
9:	/SIDSR/gcgdata/geneseq/geneseq/AA1988.DAT.*
10:	/SIDSR/gcgdata/geneseq/geneseq/AA1989.DAT.*
11:	/SIDSR/gcgdata/geneseq/geneseq/AA1990.DAT.*
12:	/SIDSR/gcgdata/geneseq/geneseq/AA1991.DAT.*
13:	/SIDSR/gcgdata/geneseq/geneseq/AA1992.DAT.*
14:	/SIDSR/gcgdata/geneseq/geneseq/AA1993.DAT.*
15:	/SIDSR/gcgdata/geneseq/geneseq/AA1995.DAT.*
16:	/SIDSR/gcgdata/geneseq/geneseq/AA1996.DAT.*
17:	/SIDSR/gcgdata/geneseq/geneseq/AA1997.DAT.*
18:	/SIDSR/gcgdata/geneseq/geneseq/AA1998.DAT.*
19:	/SIDSR/gcgdata/geneseq/geneseq/AA1999.DAT.*
20:	/SIDSR/gcgdata/geneseq/geneseq/AA2000.DAT.*
21:	/SIDSR/gcgdata/geneseq/geneseq/AA2000.DAT.*
22:	/SIDSR/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	5155	100.0	1299	22	AAW24322	Human EST encoded
2	5155	100.0	1404	13	AAW26049	MSF precursor. SY
3	5155	100.0	1404	22	AAW60568	Human megakaryocyt
4	5155	100.0	1404	22	AAW29773	Human megakaryocyt
5	2956.5	57.4	902	22	AAW29778	Human MSF-derived
6	1357.5	26.3	452	16	AAW80041	Human megakaryocyt
7	1168	22.7	5179	22	AAW24516	C899P predicted am
8	981	19.0	763	21	AAW38942	Arbidopsis thalia
9	942	18.3	1664	19	AAW43106	C. thermocellum OI
10	771.5	15.0	1325	22	AAW03645	Peptide #2327 enco
11	715	13.9	763	18	AAW31852	Mycobacterium tube

12	625.5	12.1	4412	21	AAW53666	Sequence g1/101742
13	612	11.9	572	12	AAW31855	Mycobacterium tube
14	578.5	11.2	472	22	AAW60569	Bovine MSF ortholo
15	544	10.6	844	7	AAW60570	Sequence of the Fa
16	542	10.5	807	21	AAW54467	Amino acid sequenc
17	520.5	10.1	788	21	AAW54466	Bioadhesive precu
18	506.5	9.8	744	9	AAW82975	Cryptosporidium pa
19	504	9.8	1837	21	AAW11726	Human ORFX ORF995
20	496.5	9.6	2971	21	AAW41231	Human SRCAP. Homo
21	489.5	9.5	2972	22	AAW50363	Human SRCAP. Homo
22	489.5	9.5	3118	22	AAW50362	Portion of cryptos
23	489	9.5	1721	21	AAW11727	P. yoelii SSP2 ant
24	488.5	9.5	826	13	AAW26042	Peptide #2892 enco
25	488	9.5	617	22	AAW04187	Peptide #2869 enco
26	488	9.5	617	22	AAW04187	Human Muc11 polype
27	488	9.5	957	21	AAW59288	C900P predicted am
28	488	9.5	957	22	AAW24513	Cryptosporidium pa
29	488	9.5	1721	22	AAW48299	Human protein sequ
30	485	9.4	1127	22	AAW95541	Bioadhesive precu
31	476.5	9.2	652	9	AAW82974	Peptide #1317 enco
32	467.5	9.1	511	22	AAW14883	Peptide #1349 enco
33	467.5	9.1	511	22	AAW27312	Peptide #1289 enco
34	467.5	9.1	511	22	AAW02607	PRP 378. Triticum
35	450.5	8.7	378	12	AAW14160	PRP encoded by clo
36	446.5	8.7	378	12	AAW14162	C. albicans Rbt1 p
37	446.5	8.7	750	20	AAW05477	Caenorhabditis ele
38	443	8.6	2870	21	AAW95559	Caenorhabditis ele
39	443	8.6	3178	21	AAW95556	Japanese sea mussel
40	442.5	8.6	751	16	AAW80839	S. epidermidis ope
41	425.5	8.3	910	22	AAW83007	Human atrophin-1 r
42	424.5	8.2	1012	20	AAW17406	Mouse microtubule-
43	424.5	8.2	1125	21	AAW22934	Microtubule associ
44	424.5	8.2	1125	21	AAW79637	EYFP-DEVD-MAP4-BBF
45	424.5	8.2	1610	21	AAW22870	

ALIGNMENTS

RESULT 1	
AAW24322	standard; Protein; 1299 AA.
ID	AAW24322
XX	AAW24322:
AC	12-OCT-2001 (first entry)
XX	
DE	Human EST encoded protein SEQ ID NO: 1847.
XX	
KW	Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW	tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW	diagnostics; forensic test; gene mapping; genetic disorder;
KW	bioidiversity; gene therapy; nutrition.
XX	
OS	Homo sapiens.
XX	
PN	WO200154477-A2.
XX	
PD	02-AUG-2001.
XX	
FE	25-JAN-2001; 2001WO-US02687.
XX	
PR	25-JAN-2000; 2000US-0491404.
PR	17-JUL-2000; 2000US-0617746.
PR	03-AUG-2000; 2000US-0631451.
PR	15-SEP-2000; 2000US-0663870.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI	Cao Y, Drmanac RA, Zhang J, Werhman T;
XX	
DR	WPI; 2001-476164/51.

Query Match	11.3%	Score 612;	DB 18;	Length 572;
Best Local Similarity	30.68%	Pred. No. 2.7e-27;		
Matches 193; Conservative	29;	Mismatches 253;	Indels 156;	Gaps 33;

[illegible]

RESULT 15

AAP60570
ID AAP60570 standard; Protein; 844 AA.
yy

AC AAP60570;

DT 24-AUG-1991 (first entry)
yy

DE Sequence of the *Falciparum* Interspersed Repeat Antigen
DE (FIRA).

KW Malaria vaccine; antlgen; epitope.
XY

OS Plasmodium falciparum.

PN W08601802-A

PD 27-MAR-1986

PF 11-SEP-1985; 85WO-0006960.

PR 11-SEP-1984; 84AU-0007067.

PR 10-SEP-1985; 85AU-0047326.
YY

PA (HALL-) HALL INST MED RES.
VY

Kemp DJ, Anders R, Coppel RL, Brown G, Saint RB, Cowman AF, PI xv

DR WPI; 1986-094065/14.
DR N-HCPD; 20050473

XX DNA codes for 57

poly:peptide(s) having antigenicity of RESA or FIRA antigens of F

XX
PS
Niaclouro

The inventors claim a control unit

sequence coding for RNA (AA050473) or DNA (AA050473). The

antigenicity suitable for providing protective immunity against *Plasmodium falciparum* malarial infections

Sequence 844 AA:

Query match	10.08;	Score 544;	DB 7;	Length 844;
Best local similarity	26.38;	pred NO	3	80-23;

matches 243; conservative 110; Mismatches 342; Indels 228; Gaps 47;

43 MKVSIKIIIAKPINRPSLPNSDISKETSLIVN-----KETTIVET-KE 89

vaeyssltscnnenpqtatltnltcpqisalnahlpneinetits 55

QY 755 TATTOVSTTTQDTTPFKITTLKTTTLAKVYTTKTTTTEINMKPEETAKPKDRATN 814
| : : : : :
Db 638 paappappaprvratl-----ppappappapn 665
QY 815 SKATTPPKPKP-----TKAKKPTSTKKPKTMRVRKRKPTTPPKMTSTM-----PELMP 865
| : : : : :
Db 666 smaIppapdpplpIlaetpappapplpmppapplppaapdpappllingppspIap 725
QY 866 TSRIAEAMLOTTTRPMQTPNSKLVENPKSEDAGAECEPT 905
| : : : : :
Db 726 vpgapIapIpingrpfvarknsI-----gsssgdt 756
RESULT 13
AAV53666
ID AAV53666 standard; Protein: 4412 AA.
XX AAV53666;
AC
XX
XX 22-FEB-2000 (first entry)
DT
XX
XX Sequence gI/1017427/emb/CAA62189 from an alignment with protein 608.
DE
XX
XX Mechanical stress; gene therapy; protein 608; osteoporosis; bone density;
KW bone development; gI/1017427/emb/CAA62189.
KM
XX
XX Unidentified.
OS
XX
XX WO960164-A1.
PN
XX
XX 25-NOV-1999.
PD
XX
XX 14-MAY-1999; 99WO-US11066.
PF
XX
XX 15-MAY-1998; 98US-0085673.
PR
XX
XX (QDAR-) QDARK BIOTECH INC.
PA
XX
XX Elnat P, Mor O, Skallter R, Feinstein E, Faerman A;
PI
XX
XX MPI; 2000-053304/04.
DR
XX
XX Identification of stress induced genes for determining risk and
PT preventing, treating or controlling osteoporosis
PS
XX
XX Claim 32; Fig 6A-R; 308pp; English.
CC The present sequence is obtained from a clustal X alignment with
CC protein 608. Protein 608 was identified using the method of the invention
CC after subjecting rat osteoblasts to mechanical stress. Expression of the
CC 608 gene was found to be upregulated by about 3-fold in cells subjected
CC to mechanical strain. The specification describes a method for the
CC identification of genes responsive to a specific mechanical stress. The
CC method comprises applying the mechanical stress to an organism (tissue
CC or cells comprising bone cells), isolating the specific cellular
CC fractions and extracting mRNA from them, and differentially analysing the
CC mRNA in comparison with control samples. The method is used to identify
CC genes whose expression is responsive to a specific stress. The identified
CC genes are employed in determining risk associated with a physiological or
CC disease state. The risk determination methods are used for testing a
CC medicament for gene therapy. These medicaments, or genes identified by
CC the method of the invention, are used for treating, preventing or
CC controlling a physiological or disease state (especially osteoporosis or
CC bone density or other factors causing or contributing to osteoporosis or
CC its symptoms or other conditions involved in mechanical stress or its
CC lack. The methods can also be used for advancing research or studies in
CC bone development.
XX
XX Sequence 4412 AA;
SQ

Query Match

11.5%; Score 625.5; DB 21; Length 4412;

Best Local Similarity 26.2%; Pred. No. 3,5e-27;
Matches 225; Conservative 86; Mismatches 379; Indels 169; Gaps 40;
QY 99 TDGKEKTTSAKQTOSIEKT-----SAKDLPATSKVLAKPRPKAETTTKGALL----- 146
| : : : : :
Db 3515 tdfgrlIqveeehvekvrvhvievfeaevevfkpkapp-----kgeisekIip 3566
QY 147 TKEPTPTPTT-KEPATPKKEPTPTTIKSAPTTPKEP-APTTPKSAPTTPKEPATTTKE 204
| : : : : :
Db 3567 pkkppkvvvrikeppakvpevpkIvveekvrvpeepvvpkIvpevpIpkvevp----- 3621
QY 205 PATTTPKEPATPTTKEPATTTTKSAPTTPKEPATTPKK-----PATTPKEPATTPKEP 260
| : : : : :
Db 3622 -----ekvrvppakkpkeap-----ppkveapkevpkekvvpvppkIvpevp 3672
QY 261 TPTTPKEPATTTKEPATTP-----KEPATKAKKPAATTPKEPATTPKEPATTTK 313
| : : : : :
Db 3673 kaavpekIvpeaIppkpepppevfeepespsaprk-krevpvrIvpevpvpevpkekv 3731
QY 314 EPSPTTPKEPATPTTTSAPTTPKEPATTTKSAPTTPKEPSPTTPKEPATTPKEPATPT 373
| : : : : :
Db 3732 paap-pkkrevIvpyk-vpeapkevvpkekvpyv-pkkrevpplk-----vpevpkav 3782
QY 374 PKKPAPTTPKEPATTPKEPATTTTKKPAVTAKEPATTPKEPATTPKKLT-----PTT 429
| : : : : :
Db 3783 pekkvp-----ealppkpeappppevfeepaevalleppaevveapdaappqIvppknv 3838
QY 430 PEKLAFTT-----PEKAPPTTPELAFTTPELEPTT-TTPEEPATTPPKAAPTTPKEPAT- 483
| : : : : :
Db 3839 pekkapavvakkpeliIvkvvpevpkevpkekvIvlpvkk--peapavvpevpkevp 3896
QY 484 -----TTPKEP-----APTTPKEPATTPKEPATTPPKGTAAPTTLKKEPATTPKPAKEL 534
| : : : : :
Db 3897 kIvayvkkIvpevpakvpevpkIv-----Ieekpavp-----vperaeapppvevye-peeI 3947
QY 535 APTTTPKEPTSTSDKPAPTTPKGTAPTTPKEPATTPKEPATTPKGTAPTTLKEPATPT 594
| : : : : :
Db 3948 ap-----eeelapeeekIvpyvae-----eepevpvpaapeekIvIpekIv-vIkIkeapp 3999
QY 595 PKKPAKELAPTTPKGTPTSTSDKPAPTTPKKEPA-----PTTPKEPATTPPKKP--AP 645
| : : : : :
Db 4000 pkeIpekv-----IekpIkpIppppppappkvedvkeIqIkaIpkIvkkvpepqvpeke 4056
QY 646 TTPETPPPTTSEVS--TPTTPKEP-----TTIHKSPESTPELSAEPPTKALENSPKE 696
| : : : : :
Db 4057 ItpIkvpggkkrvrlIpekIpekeevvIksvrlIvrieeepkeve-----pkIle-kvkk 4111
QY 697 PGVPTTKTPAATKP-----EKTTAKDKTTERDRLRTTPETTTAAPPKMTKEATTTETTES 752
| : : : : :
Db 4112 pavp-----eppppkpvveevvltkIreklIpeklvpeIkpIaIplpaepkIpkpea----- 4164
QY 753 KITATTTQVSTTTQDTTPFKITTLKTTTLAPVTTTKTTTTEINMK-----PEETA 806
| : : : : :
Db 4165 -----evktIkpIvpeepIlaavIvpyvgkkaeakapkeea 4204
QY 807 KPKDRATNSKATTPPKPKPTPAKPKPTS--TKKPTMPVRKPKPTPTP-----RKMTST 859
| : : : : :
Db 4205 kpgvplIgvvklIpspleaeerklIrgsggkppdeapIyqIkvaypIkvkeIkdIlt 4264
QY 860 MPELNPSTRIAEAMLOTTT 878
| : : : : :
Db 4265 esetIvgsaIfeclvpspt 4283
RESULT 14
AAW31855
ID AAW31855 standard; Protein: 572 AA.
XX AAW31855;
AC
XX
XX 27-Apr-1998 (first entry)
DT
XX
XX Mycobacterium tuberculosis 55 kDa protein.
DE

```

Dh 590 spaepenterlanenttspagptlenremtanektlfpaepenterlanekttspsa 649
Oy 480 EPAPTTPKEAPPTPKBPAPPTPKETAPPTPKGTAAPTLLEKBPAP-----TPPKBPAPKE 533
Dh 650 eptengqctplanektsspaepenthegertplanenttlspaepenterlanektlfip 709
Oy 534 LAPTTKEPTSTSDKAPAPPTPKGTAAPTLLEKBPAP-----TPPKGTAAPTLK 588
Dh 710 aeptenterlanenttspagptengrtplanektlpslaepengkrctplanektss 769
Oy 589 EPAPT-----TPKKPA-PKE-----LAPTTKGPTSTSDKBPAP 621
Dh 770 spaeptheaertplanenttspaepenterlanektlgfpaepenterlanektlfip 829
Oy 622 TTPKET-----APPTPKBPAPPTPKBPAPPTPKETPTTSEVSTPTTKEPTTI 670
Dh 830 aeptenrewtanenttlspaepenthegertplanektlpsaepenthegertplanektts 889
Oy 671 HKSPDE-----STPELSAEPYPK-----ALENSPKPEGVPT-----TKTPA 706
Dh 890 saepthegertplanektlps-raepenthegerlanekalpsakptehgetvnedtlps 948
Oy 707 ATRP-----EKTTAKDKTTERDLRTTPETTTAAPKKMTKE---TATTEKTTE 751
Dh 949 saeptengerplanenttlpsstehgetlanektlpsaepenthegertpanektlp 1008
Oy 752 SKITATP-TQVSTTODTTPFKI-----TTL-----KTTTLAPK 765
Dh 1009 spaekptheemtpsanenttspvkvkptehgetkrtlanektlpspeptehgakttsanek 1068
Oy 786 VTPT-----KTTTTEIMNKPEETAKPKDRATNSKATTPKPKPK-----825
Dh 1069 ltpslakptehgetspnckltsaaestehdratsaavtlpapaepklkhakrttlah 1128
Oy 826 -----TKAPKPTSKKKPKTMPRVKRP-----KTTPTRKMTSTIMPEL 863
Dh 1129 ekmtqytekstehpektste-kctrcpekrplysekctckgkntpvpektlenlgn 1187
Oy 864 NPTSRIAEMALQTTTRPNQT 883
Dh 1188 tlteetlkavkatenpekt 1207

RESULT 12
AAW31852
ID AAW31852 standard; Protein; 763 AA.
XX
AC AAW31852;
XX
DT 27-APR-1998 (first entry)
XX
DE Mycobacterium tuberculosis 74 kDa protein.
XX
KW Tuberculosis; mycobacteria; infection; diagnosis;
KM antimycobacterial; antidiabetic; vaccine.
XX
OS Mycobacterium tuberculosis.
XX
PN WO9741352-A2.
XX
PD 06-NOV-1997.
XX
PF 18-APR-1997; 97MO-EP01973.
XX
PR 29-APR-1996; 96DE-4017184.
XX
PA (GBFA ) GBF GES BIOTECH FORSCHUNG GMBH.
PI Espilita C, Honisch C, Moreno C, Singh M;
DR WPI: 1997-549750/50.
XX N-PSDB; AAT93610.
XX

```

```

PT New DNA and related proteins or RNA derived from M. tuberculosis -
PT used for diagnosis of mycobacterial infections, monitoring
PT vaccination and development of anti-mycobacterial agents
PS Claim 5; Fig 13; 55pp; English.
XX
CC This novel 74 kDa protein is encoded by an open reading frame of
CC a Mycobacterium tuberculosis DNA fragment (see AAT93610) containing
CC polymorphic GC-rich sequences. Its amino acid sequence shows
CC a high proline content, but there is no homology to any known
CC proline-rich antigens of mycobacteria. Novel M. tuberculosis
CC proteins (see AAW31851-57) are claimed. These can be produced as
CC recombinant proteins, especially in bacterial, yeast, fungal or
CC higher eukaryote host cells, and used for diagnosing tuberculosis
CC and other mycobacterial infections in humans or animals. The
CC claimed proteins can also be used for epidemiological studies, for
CC monitoring vaccination, and for the development of vaccines and
CC anti-mycobacterial drugs.
SQ
Sequence 763 AA:

Query Match 13.2%; Score 715; DB 18; Length 763;
Best local Similarity 28.2%; Pred No.5 4e-33;
Matches 248; Conservative 39; Mismatches 361; Indels 233; Gaps 39;

Oy 132 PTPKAETTKGPAALTTPKEPTPTPKEPASTPKETPTTIKSAPPTPKBPAPTTTKSAP 191
Dh 3 PVP-----apalalplppapapaeakskppffpap-----pppccmlvsaap 46
Oy 192 TTPKEBPATTTKBPAPTTTKBPAPTTTKBPAPTTTSAPTTKEBPAPTTPKBPAPTTKE 251
Dh 47 pcp--pap-----pappkkskafipvpapaparelappilp--pap-----peapre 90
Oy 252 PAPTPEKEPTPTT---PKEPATTKEBPAPTTPK--BPATPAK--KAPPTKEBPAPTTPK 305
Dh 91 srpalppcpvppvnlppdppeapapvpvpapnspfpfpfpapkvfpapvp--pvpnspp 148
Oy 306 EP--APTTPKEBPPTPKBPAPTTTSAPTTKEBPATTTKSAPTTKESPTTKEBPAP 363
Dh 149 fppfpaaalnppap-----papplanspplppapfpagct--ppaapvpvpaapkskpa 201
Oy 364 TTPKEBPATTPKKBPAPTTTKBPAPTTTKBPAP---TTTTKKAPAPTPKBPAPTTPEKTA 419
Dh 202 spprppap-----pmpatfimefpplpvpvppdpiskelcpapapppilppvppilp 256
Oy 420 TTPKTLPTTPEKLPAP-----TTPEKPAPTPEELAPTTPEEPTPTTPEBPAPTT-- 469
Dh 257 pvpnkilppap--ppvavaavlvapcpplpplpnhnpapapvpvppvlaplpnshp 313
Oy 470 ---PKAAAPNTPKBPAP-----
Dh 314 pappsaavpvgvnlaplpisgrvsvwksftlslfcrcvsgavlagaalnpsrpspl 373
Oy 484 ---TTPKEBPATTPKEBPAPTTKEETA---PTTPKGTGA---PTTLKEBPAPTT--KKPAPKE 533
Dh 374 tctlpalpapilppilppilpnlntavppilppvltalappilppilpilsqvgvpap-- 431
Oy 534 LAPTTTKEPTSTSDKAPAPTTPKGTAAPTLLEKBPAP-----PTTPKEBPAPTTPKGTAAPTL 587
Dh 432 --pippgkpwltlpplapappppek-lvpvlpipgscppseknpapapppppkssapalp 488
Oy 588 KEPAPTTP---KKPAPKELAPTTTGP--TSTSDKBPAPTTKET---APTTPKEBPAPTT 640
Dh 489 appapssmpsaavrvpssplppapppaapraasmpalpappspapactlcpplpsppapnsp 548
Oy 641 KKBPAPTTPEPTPTTSEVSTPTTKEPTTIHNSP---DESTPELSAEPYPKA--LENSP 694
Dh 549 --papppapilppklis--anppcpvppapnppppapppappelpappdpdpilppvansp 604
Oy 695 KPPGVPTTKTPAATKEEMTTAKDKTTERDLRTTPTETTTAAPKKMTKEATTTTEKTESKI 754
Dh 605 --pappappapsalpfvnpa-----pplcpaapk-----srpal 637

```



```

FT Domain 409..565
FT /note="cohesin type II domain"
FT Domain 607..763
FT /note="cohesin type II domain"
PN FR2748479-A1.
XX 14-NOV-1997.
PD 14-NOV-1997.
XX 10-MAY-1996; 96FR-0005854.
XX 10-MAY-1996; 96FR-0005854.
XX 10-MAY-1996; 96FR-0005854.
XX (INSP ) INST PASTEUR.
XX Beguin P, Leibovitz E;
XX WPI: 1998-011569/02.
XX N-PSDB: AAT86623.
XX Cellulase proteins with cohesin or dockerin type II domains - useful
XX for potentializing the activity of multiprotein enzyme complexes
XX Claim 7; Page 31-39; 60pp; French.
XX
CC Multimeric protein, especially enzymatic, complexes are held together
CC by protein-protein interactions between domains designated dockerins
CC and cohesins, which are found on the catalytic and scaffold subunits
CC respectively. An example of such a complex is the cellulose degrading
CC protein complex from Clostridium thermocellum, known as the cellosome.
CC This complex comprises around 15 proteins including endoglucanases,
CC cellobiohydrolases, hemicellulases, e.g. xylanases or lichenases, which
CC interact with a central "scaffold" protein designated the cellosome
CC integrating protein (CipA; see AAW43108). The catalytic subunits
CC interact with the Cip subunit via conserved 23 amino acid dockerin
CC domains. Cip has been shown to contain 9 copies of a cohesin domain.
CC The invention relates to the isolation of proteins binding to a novel
CC dockerin type domain found in the C-terminal portion of Cip. The new
CC I domain is designated a type II dockerin domain (as compared to the type
CC I domain found on the catalytic subunits of the cellosome). The type
CC II dockerin domain has some sequence similarity to the type I dockerins
CC but is unable to bind type I cohesin domains.
CC The sequence presented here is an example of a protein which binds
CC the novel type II dockerin domain and is the product of the oipb gene.
CC The protein contains 4 type II cohesin domains in the N-terminal portion
CC of which the first domain (amino acid residues 28-192) is thought to
CC bind CipA. The novel type II dockerin and cohesin domains can be used
CC in complexes, especially enzyme complexes, to potentiate their catalytic
CC actions in a synergistic manner.
XX
SQ Sequence 1664 AA;

```

```

Query Match 17.4%; Score 942; DB 19; Length 1664;
Best Local Similarity 33.0%; Pred. No. 1.8e-45;
Matches 276; Conservative 83; Mismatches 312; Indels 166; Gaps 43;

```

```

QY 128 VLAKPTP-KAETTKGPAUTTKPEPTTKPEASTTKPEPTTKSAPTTKPEAPT 186
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 758 vviqpepikaasdepltdpsdepcps-----depcps-----depltdpsd 804
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 187 TKAPPTPKPEAPTPTTKPEAPTTPKPEA---PTTKPEAPTPTTKSAPTTKPEAPT 242
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 805 eptpsdtpseepipltcdpsdepltdpsdepltdpsdepltdpsdepltdp 864
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 243 KPAPTPKPEAPTPTTKPEAPTPTTKPEAPT-TKPEAPTAPKKAAPTTPKPEAP 301
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 865 sdepltdpsdepcps--depcps--depltdpsdepltdpsdepltdpsdepltdp 919
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 302 TTPKPEA--PTTKPEAPT-TKPEAPTPTTKS-APTTPKPEAPTPTTKSAPTTPKPEAPT 357
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 920 sdepltdpsdepltdpsdepltdpsdepltdpsdepltdpsdepltdpsdepltdp 968

```

```

QY 358 TKPEAPTTPKPEAPTPTTKKPAPTTKPEAPT-TKPEAPTPTTKKPAPTKPEAPTTPKE 416
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 969 sdepcps--depcps-----depltdpsdepltdpsdepltdpsdepltdpsd--- 1019
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 417 TAPPTPKKLTPTTEKLAPTTKPEKPAPTPEELAPTTPEEPTPT-TPEEAPPTPKKAAP 475
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1020 -----epltdpsde--pltdpsdepltdps-----epltdpsdepltdpsdepltdp 1064
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 476 NTPKPEAPTTPKPEA--PTTKPEAPT-TKPEAPTPTTKGTAPTTLKPEAPTTPKPEAPK 532
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1065 tpsdepcpsdepltdpsdepltdpsdepltdpsdepltdpsdepltdpsdepcps--depcps 1122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 533 ELAPTTKEPT-STTSKPAPTTKGTAPTTKKPAPTTKPEAPTTPKPEAPTPTTKKEPA 591
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1123 d-epltdpsdepltdpsdepltdpsdepltdpsdepltdpsdepcps--depcpsdepltd 1175
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 592 PT-TPKKPAKPELAPTTKGPTSTTSKPAPT--TKPEAPTPTTKPEAPT-TPKKPAPT 646
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1176 psetpseep-----lptldpsdepltdpsdepltdpsd-epltdpsdepltdpsdepltd 1229
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 647 TPERPPTTSEVSTPTTKKEPTTIHKSPDSTPELSAEPKALENSPKEGVPTTKTPA 706
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1230 dtpsdpsd--epltdpsdepltdpsdepltdpsdepltdpsdepltdpsd--lptldpsd 1277
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 707 ATKPEMTTAKDKTTERDLTRPTETTTAARPKMTKETTESKITTTQVSTTT 766
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1278 eptpsdepltdpsdepltdpsdepltdpsdepltdpsdepltdpsdepltdpsdps 1337
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 767 QDTPTFKITTLKTTTAPKVVTTTKTITTEIMNKPEETAKPKDRATNSKATPKPKQ-KP 825
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1338 depltdpsdepltdpsdepltdpsdepltdpsdepltdpsdepltdpsdepltdpsd 1372
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 826 TKAPKKTSTKTKMPRVKPKTTPPKMTS-----TMEELNPTSRIA 870
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1373 seepdepltdpsdepltdpsdepltdpsdepltdpsdepltdpsdepltdpsd--- 1424
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 871 EAMLGTTTRPNOTPNSKLVEVNPKSEDPAGAEGETHMLLRPHVFMPEVTYPMDDIYPT 927
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1425 -----kptspaptele-epltdpsdvgalgsenraylrgy-----pdgsftrp 1465
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 10
AAB60569
ID AAB60569 standard; Protein; 472 AA.
XX
AC AAB60569;
XX
DT 27-APR-2001 (first entry)
XX
DE Bovine MSF orthologue, superficial zone protein (SZP).
XX
KW Bovine; GACP protein; campodactylly-arthropathy-coxa vara-pericarditis;
KW superficial zone protein; SZP; MSF orthologue; synovial lubricant;
KW osteoarthritis; joint lubrication; osteopontin; anilarthritis.
OS Bos taurus.
XX
WO200107068-A1.
XX
PD 01-FEB-2001.
XX
PF 21-JUL-2000; 2000WO-US20002.
XX
PR 23-JUL-1999; 99US-0145328.
PR 19-JUL-2000; 2000US-0145328.
PA (UYCA-) UNIV CASE WESTERN RESERVE.
PI Warman ML;
XX
DR WPI: 2001-182721/18.
XX
PT New composition comprising the campodactylly-arthropathy-coxa

```

PR	28-JUN-1999;	9905-0145551
PR	02-AUG-1999;	9905-0146386
PR	02-AUG-1999;	9905-0146388
PR	02-AUG-1999;	9905-0146389
PR	03-AUG-1999;	9905-0147038
PR	04-AUG-1999;	9905-0147204
PR	04-AUG-1999;	9905-0147302
PR	05-AUG-1999;	9905-0147260
PR	05-AUG-1999;	9905-0147192
PR	11-AUG-1999;	9905-0148319
PR	12-AUG-1999;	9905-0148341
PR	13-AUG-1999;	9905-0148565
PR	16-AUG-1999;	9905-0148664
PR	16-AUG-1999;	9905-0149368
PR	17-AUG-1999;	9905-0149171
PR	18-AUG-1999;	9905-0149426
PR	20-AUG-1999;	9905-0149722
PR	20-AUG-1999;	9905-0149723
PR	20-AUG-1999;	9905-0149929
PR	23-AUG-1999;	9905-0149902
PR	23-AUG-1999;	9905-0149930
PR	25-AUG-1999;	9905-0150566
PR	26-AUG-1999;	9905-0150884
PR	27-AUG-1999;	9905-0151065
PR	27-AUG-1999;	9905-0151066
PR	27-AUG-1999;	9905-0151080
PR	30-AUG-1999;	9905-0151303
PR	31-AUG-1999;	9905-0151348
PR	01-SEP-1999;	9905-0151930
PR	07-SEP-1999;	9905-0152363
PR	10-SEP-1999;	9905-0153070
PR	13-SEP-1999;	9905-0153758
PR	15-SEP-1999;	9905-0154018
PR	16-SEP-1999;	9905-0154038
PR	20-SEP-1999;	9905-0154779
PR	22-SEP-1999;	9905-0155139
PR	23-SEP-1999;	9905-0155139
PR	28-SEP-1999;	9905-0155486
PR	28-SEP-1999;	9905-0155659
PR	29-SEP-1999;	9905-0156458
PR	29-SEP-1999;	9905-0156596
PR	05-OCT-1999;	9905-0157111
PR	05-OCT-1999;	9905-0157153
PR	06-OCT-1999;	9905-0157865
PR	07-OCT-1999;	9905-0158029
PR	08-OCT-1999;	9905-0158232
PR	12-OCT-1999;	9905-0158369
PR	13-OCT-1999;	9905-0158293
PR	13-OCT-1999;	9905-0158294
PR	13-OCT-1999;	9905-0158295
PR	14-OCT-1999;	9905-0159329
PR	14-OCT-1999;	9905-0159330
PR	14-OCT-1999;	9905-0159531
PR	14-OCT-1999;	9905-0159537
PR	21-OCT-1999;	9905-0160768
PR	21-OCT-1999;	9905-0160770
PR	21-OCT-1999;	9905-0160814
PR	21-OCT-1999;	9905-0160815
PR	22-OCT-1999;	9905-0160980
PR	22-OCT-1999;	9905-0160981
PR	22-OCT-1999;	9905-0160984
PR	25-OCT-1999;	9905-0161405
PR	25-OCT-1999;	9905-0161406
PR	26-OCT-1999;	9905-0161359

PR	26-OCT-1999;	99US-0161360
PR	26-OCT-1999;	99US-0161361
PR	28-OCT-1999;	99US-0161920
PR	28-OCT-1999;	99US-0161922
PR	28-OCT-1999;	99US-0161993
PR	29-OCT-1999;	99US-0162142

Query Match	18.1%	Score 981	DB 21	Length 763
Best Local Similarity	38.5%	Pred No. 5.1e+48		
Matches 226	Conservative 24	Mismatches 293	Indels 44	Gaps 11

```

0Y 137 ETTTKGPAALTTKPEPTPTTKKEPASTTKPEPTPTTKSATTTKEBPATTKKAPPTTKE 196
Db 126 qfgaguruyllpppprlpppppcvtlcp-eps-----pppvtc---pqdpt 170
0Y 197 PAPTTTKBPATTTKBPATTTTKBPATTT--TKSATPTTKBPATTPKKKAPTTKBPAP 254
Db 171 lprltprpctppgpaupctvclvtgppcllppqgprlrlpdelppnnpqgprllp----p 226
0Y 255 TTPEKPEPTTTKBPATTT--KEBPATTTKEBPATPAKKKAPATTTKBPATTTKBPATTTK 313
Db 227 ntpcrlptltpcprpnrcpnsprnlpdelprnlpqgprntprntprcrlptltpcrlptlp 286
0Y 314 EPSPTTKBPAPATTTKKSAPTTTKBPAPTTTKSATTTKEPSPTTTKBPATTTKBPAPAT 373
Db 287 tcsprntlptscrlptclprnlpqpss-----pnlpcprpnlpqgprntprntpcrlpnt 340
0Y 374 PKKAPATTPKBPATTPKBPATTTKKKAPAPAKBPAPATTPKBPATTPKKKlpttPEKL 433
Db 341 ppgltpnrlpccrlptltpcrlptlrlpccprnrltpqgspnrlpccprnrltpqgprnrlpcc 400
0Y 434 ABTTKEKAPATTPPELAPTTPEEPPTTTKEBPATTTKEAAPNTTKBPATTTKBPATTT 493
Db 401 lprntpqgprnrlpccrlptltpcrlptltpcrlntlpqgspnrlpccprnrltpqgprnt 460
0Y 494 PKBPATTTKEAPATTPKGTATTTKEBPATTPKKKAPAKELAPATTTKEPSTSDKAPAT 553
Db 461 ptkp-rltpcprnrlpccrlptltpcrlpnt-----pvgprpdltpcrlpntprcrlpn 514
0Y 554 TPKGATPTTKBPAPATTTKBPATTPKGTATTTKEBPATTPKKKAPAKELAPTTTKGPTS 613
Db 515 ltpqgprnqgprlprlprntprcrlpccrltpnspcrlrlprntprcrlpntlpqgspcl 574
0Y 614 TTSDKP-----APPTKEAPATTPKBPAPTTKPKKAPATTPETPRPTTSVSPPTTKE 666
Db 575 lprntprqglprlprntprcrlpntprcrlpntlpqgprlprntprntprntprntprntpr 634
0Y 667 P---TTTHKSPDESPTPELSAETPKALESPKPEPVPTTKTPAATKP 710
Db 635 ppcltllpspsptcspnmspspppksgppppprlfgpppprctgpc 681

```

	RESULT	9
AAM43106	ID	AAM43106 standard; Protein; 1664 AA.
XX		
XX	AC	AAM43106;
XX		
DT		16-OCT-1998 (first entry)
XX		
DE		C. thermocellum OLPB protein.
XX		
KW		Multimer; enzyme; complex; protein-protein interaction; dockerin domain
KM		cohesin domain; catalytic subunit; scaffold subunit; SDBA; synergistic;
KM		cellulosome integrating protein; scaffoldin dockerin binding protein.
XX		
OS		Clostridium thermocellum.
XX		
FH	Key	Location/Qualifiers
FT	Domain	28..192
FT		/note="cohesin type II domain"
FT	Domain	207..363
FT		/note="cohesin type II domain"

PR	19-MAY-1999:	99US-0134641
PR	20-MAY-1999:	99US-0135124
PR	21-MAY-1999:	99US-0135153
PR	24-MAY-1999:	99US-0135629
PR	25-MAY-1999:	99US-0136021
PR	27-MAY-1999:	99US-0136392
PR	28-MAY-1999:	99US-0136582
PR	01-JUN-1999:	99US-0137222
PR	03-JUN-1999:	99US-0137528
PR	04-JUN-1999:	99US-0137502
PR	07-JUN-1999:	99US-0137724
PR	08-JUN-1999:	99US-0138094
PR	10-JUN-1999:	99US-0138540
PR	10-JUN-1999:	99US-0138847
PR	14-JUN-1999:	99US-0139119
PR	16-JUN-1999:	99US-0139453
PR	17-JUN-1999:	99US-0139453
PR	18-JUN-1999:	99US-0139454
PR	18-JUN-1999:	99US-0139455
PR	18-JUN-1999:	99US-0139456

PR 18-MAY-1999; 99US-0134768

PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.

CC particular, one of the surfaces is pericardial tissue: DNA encoding a
 CC tribonectin may be used in gene therapy. The present sequence represents
 CC a substantial portion of a human MSF-derived tribonectin.
 XX

SO Sequence 902 AA:

Query Match 59.4%; Score 3217.5; DB 22; Length 902;
 Best Local Similarity 73.4%; Pred. No. 27e-174;
 Matches 744; Conservative 28; Mismatches 80; Indels 161; Gaps 65;

```

OY 1 VKDNKKNRRTKKRPKPPVVDGSGLDNGDKVTTPTDSTTOHNNKVTSPKITTAKPIN 60
   |||
DB 1 vkdnknrtkkrrpkrpvpvdeagsgldngdfkvtptdsttcqnhkvstspkittakpin 60
OY 61 PRPSLPNSDTSKENSILYNNKEETVETETTTNNKOTSDGKEKTSKKEQSIKTSK 120
   |||
DB 61 prpslpnsdtskenskylvnnketvettetttnnkotsgdkekltsakelsktsak 120
OY 121 DLAPTSKVLAKRPKAEITTKGPAITTPKEPPTTPKEPASTTPKEPPTTKSAPTTPK 180
   |||
DB 121 dlaptskvlakrpkaeitttkgpaittpkeppttpkepasttpkeppttksap 180
OY 181 EPAPTTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKSAPTTPKEPAPTT 240
   |||
DB 173 epapttksaptttkkepapttkkepapttkkepapttkkepapttkkepapttk 228
OY 241 PKKRAPTPKEPAPTTTPKEPPTTPKEPAPTTTPKEPAPTTTPKAPTAAPKRAPTPKEPA 300
   |||
DB 229 -kepapttkkepapttkkepapttkkepapttkkepapttkkepapttkkepapttk 281
OY 301 PTPPEAPPTTPKEPPTTPKEPAPTTTPKSAPTTKKEAPPTTPKSAPTTPKEPPTTPKE 360
   |||
DB 282 ptt-kepapttkkepapttkkepapttkkepapttkkepapttkkepapttkkepapttk 335
OY 361 PAPPTPKAPAPTTTPKRAPATTPKRAPATTPKRAPATTPKRAPATTPKRAPATTPK 420
   |||
DB 336 pappttkrapattpkrapattpkrapattpkrapattpkrapattpkrapattpkr 388
OY 421 TPKLITPTPEKLAPTTPPEKAPPTTPPELAAPTTPEEPPTTPPEAPATTPKAAAPNTPK 480
   |||
DB 389 t-kepapttkkepapttkkepapttkkepapttkkepapttkkepapttkkepapttk 440
OY 481 PAPPTPKAPAPTTTPKRAPATTPKRAPATTPKRAPATTPKRAPATTPKRAPATTPK 540
   |||
DB 441 pappttkrapattpkrapattpkrapattpkrapattpkrapattpkrapattpkr 488
OY 541 EPTSTSDKRAPPTTPKGTAPTPKEPAPTTTPKRAPATTPKGTAPTPKRAPATTPK 600
   |||
DB 489 epapttkrapattpkrapattpkrapattpkrapattpkrapattpkrapattpkr 541
OY 601 KEAAPTTTKAPPTSTSDKRAPPTTPKRAPATTPKRAPATTPKRAPATTPKRAPATTPK 660
   |||
DB 542 -ttkepapttkkepapttkkepapttkkepapttkkepapttkkepapttkkepapttk 589
OY 661 PTTTKEPTTHKSPDSTPELSAEPKPALENSPKRPGVPTTKPATRPEMTTAKDKT 720
   |||
DB 590 -pttkkepapttkkepapttkkepapttkkepapttkkepapttkkepapttkkepapttk 636
OY 721 TERDLRTPEPTTAAPKMTKETATTEKTTESKITTATTQVSTTTODTTPKITTAKTT 780
   |||
DB 637 -papttkkepapttkkepapttkkepapttkkepapttkkepapttkkepapttkkepapttk 667
OY 781 TLAPKVTTKKTTTTELNKKPEETAKPRDRATNSKATTPKPKQPTKAPKPKTSTSKPKPT 840
   |||
DB 668 -ttkepapttkkepapttkkepapttkkepapttkkepapttkkepapttkkepapttk 686
OY 841 MPFRVRRKPTTPPKMTSTMPELNPTSRITAEAMLOTTRPNTQSKLVENPKSEDDG 900
   |||
DB 687 -aptrprrkpttpkrmtstmpelnptsrtaeamlottrpntqsklvnpksekedag 739
OY 901 AGETPHMLLRHVMPREVTPEMDYLRVVPNGIILNMLSDETICNGKRVVDGTLTLRN 960
   |||

```

DB 740 agetphmlrrhvfmprevtpmdylrvvpngiilnmlsdetlcnkgkpvdgltlrn 799
 OY 961 GTLVAFRGHYFWMLSPSPSPAPRRITTEWGISPIDTVPTRCNCEGKTFE 1013
 DB 800 gtlvafghyfwmlspspsparrltevwgispidvtvtrcncegkttffk 852

RESULT 6
 AAR80041
 ID AAR80041 standard; Protein; 452 AA.
 AC AAR80041:
 XX 10-APR-1996 (first entry)
 DE Human megakaryocytopoietin protein.
 XX
 KW Human: megakaryocytopoietin; wheat germ agglutinin; heparin;
 KW megakaryocyte; aplastic anaemia; thrombocytopenia; hematocytopenia;
 XX multipotential stem cell.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 393..396
 FT /note= "unspecified amino acids"
 FT Misc-difference 444..446
 FT /note= "unspecified amino acids"
 PN W09523861-A1.
 XX
 PD 08-SEP-1995.
 XX
 PE 06-MAR-1995; 95WO-CN00015.
 XX
 PR 04-MAR-1994; 94CN-0112066.
 XX
 PA (SHAN-) SHANGHAI BEITI BIOTECHNOLOGY CO LTD.
 XX
 PI Gu X, Han Z, Shen Q;
 XX
 DR WPI. 1995-320576/41.
 DR N-PSDB; AAT04546.
 XX
 PT New haematopoietic cell growth factor - used for treating
 PT thrombocytopenia and hematocytopenia
 PS
 XX
 PS Example: Page 23; 36pp; Chinese.
 XX
 CC This sequence represents the human megakaryocytopoietin (MPO) protein.
 CC This sequence was purified using a carrier which can couple wheat germ
 CC agglutinin and heparin to separate MPO. Fragments of this sequence (see
 CC AAR80039 and AAR80040) were used to produce the amplification primers
 CC shown in AAT04544 and AAT04545. The fragments amplified by these primers
 CC can then be used as probes to screen human cDNA libraries for MPO cDNA.
 CC The MPO cDNA can then be inserted into a plasmid which is used to
 CC transform cells to produce MPO. The MPO sequence is capable of promoting
 CC colony formation of megakaryocytes, enlarging the size of megakaryocytes
 CC and stimulating the proliferation of multipotential stem cells. The
 CC factor may be used for treating thrombocytopenia and hematocytopenia.
 CC The purification method can be used to isolate MPO from human urine or
 CC sera of patients with aplastic anaemia, and from animal blood or urine
 CC by radiation exposing the animals to induce aplastic anaemia.
 XX
 SO Sequence 452 AA;

Query Match 25.1%; Score 1357.5; DB 16; Length 452;
 Best Local Similarity 88.3%; Pred. No. 1.6e-69;
 Matches 263; Conservative 3; Mismatches 21; Indels 11; Gaps 1;
 OY 1 VKDNKKNRRTKKRPKPPVVDGSGLDNGDFVTTPTDSTTOHNNKVTSPKITTAKPIN 60
 |||

measuring the amount of MSF or its fragment in a biological sample of a mammal, wherein an increased amount of MSF compared to a control indicates the presence of or predisposition to developing osteoarthritis. The tribonectin and DNA encoding it are useful in the treatment of osteoarthritis, where they may be used for lubricating mammalian joints, such as articulating joints of humans, dogs or horses. The tribonectin, when formulated as a membrane, foam, gel or fibre, is useful for inhibiting adhesion between two surfaces such as the injured tissues of a mammal, where the injury is caused by a surgical insertion or trauma, or an artificial device e.g., an orthopaedic implant. In particular, one of the surfaces is pericardial tissue. DNA encoding a tribonectin may be used in gene therapy. The present sequence represents human MSF.

Sequence 1404 AA;

Query Match 100.0%; Score 5416; DB 22; Length 1404;
Best Local Similarity 100.0%; Pred. No. 2,66-298;
Matches 1013; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 VKDNKKNRTKKKPPKPPVVDAGSGLDNGDFKVTTPDSTTQHNKVSPTKITTAKPIN 60
DB 200 vkdnknrtkkkppkppvvdagsgldngdfkvttpdsttqhnkvsptkittakpin 259
QY 61 PRPSLPNSDTSKETSILVKNKETVTETKETTNNKOTSDGKEKTTSAKETOSIEKTSK 120
DB 260 prpslpnsdtsketsilvknketvtetkettnnkotstdgkektttsaketosiekttsak 319
QY 121 DLAPTSTKVLAKPPPKAETTITGKAPALTTPKPPPTTPKEPASTTPKPEPTTISAPTTPK 180
DB 320 dlaptstkvlakpppkatettitgkapalttpkpppttpkepasttpkpepttisapttk 379
QY 181 EPAPTTSKAPTPPKPEAPPTTTPKEAPPTTTPKEAPPTTTPKEAPPTTTPKSAFTTPKEAPPT 240
DB 380 epapttskaptppkpeapptttpkeapptttpkeapptttpkeapptttpkcappttk 439
QY 241 PKKPAPTTPKEAPPTTTPKEAPPTTTPKEAPPTTTPKEAPPTTTPKEAPPTTTPKEAPPT 300
DB 440 pkkpapttppkeapptttpkeapptttpkeapptttpkeapptttpkeapptttpkeappt 499
QY 301 PPTTKREAPPTTTPKEAPPTTTPKEAPPTTTPKSAFTTPKSAFTTPKESPTTKE 360
DB 500 ppttkreapptttpkeapptttpkeapptttpkcappttkcappttkcappttk 559
QY 361 PAPTPPKPEAPPTTTPKAPPTTTPKEAPPTTTPKAPPTTTPKAPPTTTPKAPPTTTPKAPPT 420
DB 560 paptppkpeapptttpkcapptttpkeapptttpkcappttkpaptcaptkpaptcap 619
QY 421 TPKKLTPTTPPEKLAFTTPPEKAPPTTPEELAPTPPEEPPTPEEPAPPTTTPKAAAPNTPK 480
DB 620 tpkkltpttppeklafttppekapttpeelaptppeepptpeepappttppkaapnptke 679
QY 481 PAPTPPKPEAPPTTTPKAPPTTTPKAPPTTTPKAPPTTTPKAPPTTTPKAPPTTTPKAPPT 540
DB 680 paptppkpeapptttpkcapptttpkcapptttpkcapptttpkcapptttpkcapptt 739
QY 541 EPTSTTSDKAPPTTTPKAPPTTTPKAPPTTTPKAPPTTTPKAPPTTTPKAPPTTTPKAPPT 600
DB 740 eptsttsdkapptttpkcapptttpkcapptttpkcapptttpkcapptttpkcappt 799
QY 601 KELAPPTTTPGPTSTSDKAPPTTTPKAPPTTTPKAPPTTTPKAPPTTTPKAPPTTTPKAPPT 660
DB 800 kelapptttpgptstsdkapptttpkcapptttpkcapptttpkcapptttpkcapptt 859
QY 661 PPTTKPEPTTTPKSPDESTPELSAEPPTPKALENSPKPEGVPTTKTTPAATPEMTTAKOKT 720
DB 860 ppttkpeptttpkspdestpelsaepptpkalenspkpegvpttktpaatpemttaokt 919
QY 721 TERDLRPTTPPTTTPAARPKTKETATTEKTTESKITTATTOVSTTQODTPPKITTLTKT 780
DB 920 terdlrpttpptttpaarpktketattektteskittattovtsttqodtppkittlkt 979
```

```
QY 781 TLAPKYTTTKTTTITTEIMNKPEETAKPKDRATNSKATTPPKPOKPTKAPKPEPTSTRKEPT 840
DB 980 tlapytttktttittetimnkpeetapkdratnkattppkpokptkapkpeptstrkept 1039
QY 841 MPVRKPKTTPTRPKKMTSTWPELNPPTSRIAEMALQTTTRPNOTPNSKLYEVNPKSDDAG 900
DB 1040 mpvrpktptrpkkmstwpeelnpptrsriaemalqtttrpnotpnsklyevnpkseddag 1099
QY 901 AEGSTPMMLLRPHVMEVTPDDMDYLPRVNPNGIITINPMLSDETNICNGKPYDGLTTLN 960
DB 1100 aegstpmmlrphvmevtpddmdylprvnpngiitinpmlsdetnicngkpydglttltn 1159
QY 961 GEIVAFGHYFMMLSPSPSPSPARRITEVWGIPSPIDVTFRNCBGKTPFEK 1013
DB 1160 geivafghyfmmlspspsparritevwgipspidvtfrncbgktpfeek 1212

RESULT 5
AAB29778
ID AAB29778 standard; Protein: 902 AA.
AC
XX AAB29778:
XX
XX 28-FEB-2001 (first entry)
DT
XX
XX Human MSF-derived tribonectin.
DE
XX Human tribonectin; MSF; megakaryocyte stimulating factor;
KM alternative splicing; joint boundary lubricant; O-linked oligosaccharide;
KM osteoarthritis; tribosupplementation; tissue adhesion inhibition;
KM friction coefficient reduction; gene therapy; antiarthritic;
KM osteopathic.
XX
XX Homo sapiens.
OS
XX WO200064930-A2.
PN
XX 02-NOV-2000.
PD
XX 24-APR-2000; 2000WO-US10953.
PE
XX 23-APR-1999; 99US-0298970.
PR
XX (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.
PA
XX Jay GD;
PI
XX WPI; 2001-024673/03.
DR
XX Novel tribonectin polypeptide useful as lubricant for treating
XX osteoarthritis, comprises O-linked lubricating moiety
PT
XX Disclosure: Fig 1; 47pp; English.
PS
XX
XX The invention relates to a human tribonectin which is a product of
XX alternative splicing of the human MSF (megakaryocyte stimulating factor)
XX gene. The tribonectin has at least one O-linked oligosaccharide
XX lubricating moiety and has a polypeptide sequence comprising 1-76
XX repeats of a motif having at least 50% identity to the sequence KEPAPTT
XX (AAB29774). The invention also relates to a nucleic acid encoding a
XX human MSF-derived tribonectin; a biocompatible composition comprising a
XX human tribonectin for inhibiting tissue adhesion formation; and a method
XX of diagnosing osteoarthritis or a predisposition to osteoarthritis by
XX measuring the amount of MSF or its fragment in a biological sample of a
XX mammal, wherein an increased amount of MSF compared to a control
XX indicates the presence of or predisposition to developing
XX osteoarthritis. The tribonectin and DNA encoding it are useful in the
XX treatment of osteoarthritis, where they may be used for lubricating
XX mammalian joints, such as articulating joints of humans, dogs or horses.
XX The tribonectin, when formulated as a membrane, foam, gel or fibre, is
XX useful for inhibiting adhesion between two surfaces such as the injured
XX tissues of a mammal, where the injury is caused by a surgical insertion
XX or trauma, or an artificial device e.g., an orthopaedic implant. In
```

XX Example 1: Page -: 34pp; English.
 PS The invention relates to a method of treating osteoarthritis via the
 XX administration of a composition comprising the campodactylly-arthropathy-
 CC coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.
 CC The composition may further comprise a local anaesthetic. The composition
 CC of the invention may be administered via intra-articular or intravenous
 CC injection. The human CACP protein is identified in the invention as
 CC being megakaryocyte stimulating factor (MSF). The gene encoding
 CC CACP protein (MSF) is located on chromosome 1q25-31, and mutations in
 CC this gene are responsible for the heritable disorder campodactylly-
 CC arthropathy-coxa vara-pericarditis, in which patients have synovial
 CC hyperplasia without evidence of inflammation. CACP protein (MSF)
 CC acts as a synovium lubricant, and can be used to lubricate tissue and
 CC joints in the treatment of osteoarthritis. The composition may be
 CC applied to reduce the symptoms of osteoarthritis (e.g., joint pain,
 CC loss of range of movement or joint damage). The present sequence
 CC represents human megakaryocyte stimulating factor (MSF, CACP protein).
 CC Note: This sequence is not given in its entirety in figure 4 of the
 CC specification, although a Genbank accession number was given. This
 CC sequence was therefore obtained from GenBank (U70316).
 XX
 XX Sequence 1404 AA:

Query Match 100.0%; Score 5416; DB 22; Length 1404;
 Best Local Similarity 100.0%; Pred. No. 2,6e-298;
 Matches 1013; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKDKNNRTRKKKPPKPPVVDGSLDNGDFKTPDSTTQNNKSTSKITTAAPIN 60
 DB 200 vkdknnrtrkkkppkppvvdgslndgdfktpdsttqnnkstsksittakpin 259
 QY 61 PRSLPNSDTSKSTSLTVNKEETVETKETTNNKQSTGKKEKTAQSIEKTSAS 120
 DB 260 prslpnsdtskstsiltvnkettvetkettnnkqstgkkttsaketsieksas 319
 QY 121 DLATSVLVAKPPPKAATTGKALTPPKKEPTTPPKKPASTPKKEPTPTTISAPTPK 180
 DB 320 dlatssvlvakpppkattgkaltpkpkpkepttpkkpaspkpkpkeptpttisaptpk 379
 QY 181 EPAPTTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKAPPTTSAPTTPKEPAPT 240
 DB 380 epapttksaptpkpeaptpkpeaptpkpeaptpkpeaptpkksaptpkpeaptp 439
 QY 241 PKKPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKAPPTTPKEPA 300
 DB 440 pkkpaptpkpeaptpkpeaptpkpeaptpkpeaptpkpeaptpkksaptpkpeaptp 499
 QY 301 PTTPKKEPAPTTTPKEPAPTTTPKEPAPTTTPKSAPTTPKSAPTTPKEPSPPTTKE 360
 DB 500 pttpkpeaptpkpeaptpkpeaptpkksaptpkksaptpkksaptpkpeaptpk 559
 QY 361 PAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKAPPTTTPKAPPTTPKEAPAP 420
 DB 560 papttpkpeaptpkpeaptpkpeaptpkpeaptpkpeaptpkpeaptpkpeaptp 619
 QY 421 TPKKLTPTTPKEKAPPTTPKEPAPTTTPKEEAPPTTPEEPPTTPKEEAPPTTPKAAAPMTPK 480
 DB 620 tpkkltpkpeaptpkpeaptpkpeaptpkpeaptpkpeaptpkpeaptpkpeaptpk 679
 QY 481 PAPTTPKEPAPTTTPKEPAPTTTPKETAPTTPKGAAPTTTLKEPAPTTTPKKAPELATTTTK 540
 DB 680 papttpkpeaptpkpeaptpkpeaptpkpeaptpkpeaptpkpeaptpkpeaptpk 739
 QY 541 EPTSTSDKAPATTPKGAAPTTTPKEPAPTTTPKGAAPTTTLKEPAPATTPKPPAP 600
 DB 740 eptstsdkapattpkgaaptpkpeaptpkpeaptpkpeaptpkpeaptpkpeaptpk 799
 QY 601 KETAPPTTGGPTSTSDKAPATTPKGAAPTTTPKEPAPTTTPKKAPELATTTTPPTTSEVST 660
 DB 800 ketappttggptstsdkapattpkpeaptpkpeaptpkpeaptpkpeaptpkpeaptpk 859

QY 661 PTTKKEPTTTIKSPDESTPELSAEPKPALENSKEBGPVTTTKTPAATKEMTTAKDKT 720
 DB 860 pttkkeptttikspdestpelsaepkpaalenskebgpvtttktpaatkemttakdk 919
 QY 721 TERDLRTPEPTTAAAPKMTETATTEKTESKTTATTTQVSTTODTTPFKITTLKTT 780
 DB 920 terdlrtpeptttaaapkmetattekteskttattttqvsttoddtpfkittlkt 979
 QY 781 TLADKVTTKTKTTTTEIMNKPDEETAKPKDRAATNSKATTPKPKTPAKPKPTSTKKPT 840
 DB 980 tladvtktkttttteimnkpedetapkdratnskattpkpktpakpkptstkkpt 1039
 QY 841 MPRRKKTTPTPPKMTSTPELNPSTRISIAEMLQTTTRPNQPNKSLVEVNPKSDAG 900
 DB 1040 mprvrkkttpkmtstpelnpstrisiamlqtttrpnqpnkslvevnpksedag 1099
 QY 901 AEGETPMALRPHVFMPEVTPDMQYLPRVNPNOGIITPMISDEFNTCNCKEPVGLTLRN 960
 DB 1100 aegetpmlrphvfmpevtpdmqylprvnpngiitpmlsdefntcngkpvgltlrn 1159
 QY 961 GTLVAFRGHYFWMLSPESSPARRTTEVWGISPIDVTFRNCNCEGKTFEKK 1013
 DB 1160 gltvafrghyfwmlspessparrttevwgispidvfrncncegkctfeffk 1212

RESULT 4

AA29773
 ID AAB29773 standard; Protein; 1404 AA.

AC AAB29773;
 XX

DT 28-FEB-2001 (first entry)
 XX

DE Human megakaryocyte stimulating factor (MSF), SEQ ID NO:1.
 XX

KW Human MSF; megakaryocyte stimulating factor; tribonectin;
 XX

KW osteoarthritis; tribosupplementation; tissue adhesion inhibition;
 XX

KW friction coefficient reduction; gene therapy; antiarthritic;
 XX

OS Homo sapiens.
 XX

PN WO200064930-A2.
 XX

PD 02-NOV-2000.
 XX

PF 24-APR-2000; 2000WO-US10953.
 XX

PR 23-APR-1999; 99US-0298970.
 XX

PA (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.
 XX

PI Jay GD;
 XX

DR WPI: 2001-024673/03.
 XX

PT N-PSDB; AAC81498.
 XX

PT Novel tribonectin polypeptide useful as lubricant for treating
 XX osteoarthritis, comprises O-linked lubricating moiety

XX Claim 3; Page 7; 47pp; English.

XX The invention relates to a human tribonectin which is a product of
 CC alternative splicing of the human MSF (megakaryocyte stimulating factor)
 CC gene. The tribonectin has at least one O-linked oligosaccharide
 CC lubricating moiety and has a polypeptide sequence comprising 1-76
 CC repeats of a motif having at least 50% identity to the sequence KKPAPT
 CC (AAB29774). The invention also relates to a nucleic acid encoding a
 CC human MSF-derived tribonectin; a biocompatible composition comprising a
 CC human tribonectin for inhibiting tissue adhesion formation; and a method
 CC of diagnosing osteoarthritis or a predisposition to osteoarthritis by

DR WPI: 1992-284660/34.
N-PSDB: AAQ27223.

XX New human mega-karyocyte stimulating factors - for treating
PT immune deficiencies, cancer, exposure to radiation or drugs,
XX bacterial and viral infections, etc.

PS Claim 1, 2 and 3; Fig 1; 87pp; English.

XX The sequence given is a full length translation from the megakaryocyte
CC stimulating factor (MSF) precursor. The sequence covered by exons II,
CC III and IV encodes megakaryocyte stimulating factor (MSF). This
CC sequence is modified by the addition of an N-terminal sequence encoding
CC a secretory leader, an initiating methionine preceding exon II and a
CC terminating codon following exon IV. The cDNA sequence given contains
CC sequences derived from human megakaryocyte colony stimulating factor
CC (meg-CSF). Exon I contains the initiating methionine, and encodes a
CC classical mammalian protein secretion signal sequence. The sequence
CC encoding the original meg-CSF includes exons II-IV and is thought to
CC terminate in the region between amino acid residues 134 - 147. The
CC primary transcript of this gene may be cleaved in different ways to
CC yield a family of mRNA's each encoding a different MSF protein. Exons
CC V and VI are thought to be related to the activity of the factor and
CC are also implicated in the stability, folding and processing of the
CC molecule. These exons are also thought to play a role in the observed
CC synergy of MSF with other cytokines. Exons V - XII are believed to be
CC implicated in the processing or folding of the appropriate structure of
CC the resulting factor, i.e. one or more of these exons may contain
CC sequences which direct proteolytic cleavage, adhesion, organisation of
CC the cellular matrix or extracellular matrix processing. Both naturally
CC occurring and non-naturally occurring MSF's may be characterised by
CC various combinations of alternatively spliced exons from this sequence,
CC with the exons spliced together in differing orders to form different
CC members of the MSF family.

XX Sequence 1404 AA:

Query Match 100.0%; Score 5416; DB 13; Length 1404;
Best Local Similarity 100.0%; Pred. No. 2,6e-298;
Matches 1013; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKNNKRRTRKKRPKPVYDEAGSLNDGDFVTTPDSTTHNKNVTSPIKTTAKPIN 60
DB 200 vknnkrrtrkkpkrpvvdeagsgldngdfvttpdstthnknvstpkittakpin 259
QY 61 PRPSLPNSDTSKETSILTNNKETTVEKETTNNKQTSIDGKEKTSAKETOSTEKTSAK 120
DB 260 prpslpnsdstsketsiltnnkettvetketttnkqtsidgkekttsaketsaksak 319
QY 121 DLAPTSKVLAKPPPKATTTTGGALTTTPKPEPTTPPKPEASTTPPKPEPTTTSAPTPK 180
DB 320 dlaptskvllakpppkattttggaltttpkpepttpkpeasttpkpepttttsaptpk 379
QY 181 EPAPTTKSAPTPKPEAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 240
DB 380 epapttksaptpkpeaptttkepaptttkepaptttkepaptttkepaptttkepaptt 439
QY 241 PKKPAPTTPKPEAPTTPKPEPTTPPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTT 300
DB 440 pkkpapttpkpeapttpkpepttpkpeapttpkpeapttpkpeapttpkpeapttkepa 499
QY 301 PTTTKEAPTTTKEPSPPTTPKPEAPTTTTSAPTTTKEPAPTTTKEPAPTTTKEPAPTT 360
DB 500 ptttkeaptttkepspttpkpeapttttsaptttkepaptttkepaptttkepspttke 559
QY 361 PAPTTPKEAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEAPTT 420
DB 560 papttpkeaptttkepaptttkepaptttkepaptttkepaptttkepaptttkeaptt 619
QY 421 TPKKLTPTTPPKLAPTTPEKAPPTTPEELATTTPEEPPTTPPEKAPPTTPEKAPPTTPE 480
DB 620 tpkkltpttpkklapttpekappttpeelatttpeeppttppekappttpekappttpe 679

QY 481 PAPITPKPEAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 540
DB 680 papitpkpeaptttkepaptttkepaptttkepaptttkepaptttkepaptttkepapttt 739
QY 541 EPTSTTSKRPAPTTTPKAGAPTTTPKPEAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 600
DB 740 eptsttskraptttpkagaptttpkpeaptttkepaptttkepaptttkepaptttkepaptt 799
QY 601 KELAPTTTSGPTSTSDKAPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 660
DB 800 kelaptttsgptstsdkapaptttkepaptttkepaptttkepaptttkepaptttkepaptt 859
QY 661 PTTTKEPTTHKSPESPPELSAETTPKALENSKPEPEVPTTPKPAATKPMATTTAKDKT 720
DB 860 ptttkeptthkspesppeelsaettpkalenskpepevpttpkpaatkpmatttakdktt 919
QY 721 TERDLRTPEPTTTAAPKMTKETATTTETTESKITATTTQVTSSTTTQOTTPPKITLTKTT 780
DB 920 terdlrtpeptttaapkmtketatttetteskitatttqvtsstttqottppkittlkttt 979
QY 781 TLAPVTTTTKITTTTETLNNKPEETAKPKDRTNKSATTPKPKOKPTKPKPTSKPKPT 840
DB 980 tlapvtttktitttettelnnkpeetakpkdrtnskattppkpkokptkpkptskpkpt 1039
QY 841 MPVRKRPKTTTPPKMTSNMPLNPTSRIAEAMLOTTTRPNQTPNSKLVENVPKSDEAGG 900
DB 1040 mpvrkrpktttppkmtsnmplnptsriaeamlotttrpnqtpnsklvenvpkseadagg 1099
QY 901 AAGETPHMLLRPHVEMPEVTPMDYLPVFNOCIIINPMLSEPTNINCNCPPVDGLTTLRN 960
DB 1100 aagepthmlrrphvempevtpmdylpvfnociiinpmlseptnincncppvdglttlrn 1159
QY 961 GRTIYARGHYFMWLSPESPSPARITEVWGIJSPIDVFTFCNCEGKTFEKK 1013
DB 1160 grtiyarghyfmwlspespsparitevwgijspidvftfcncegkttfeffk 1212

RESULT 3
AAB60568
ID AAB60568 standard; Protein: 1404 AA.
XX
AC AAB60568;
XX
DT 27-APR-2001 (first entry)
XX
DE Human megakaryocyte stimulating factor (MSF, CACP).
XX
KW Human; CACP protein; campodactylly-arthropathy-coxa vara-pericarditis;
KW MSF; megakaryocyte stimulating factor; synovial lubricant;
KW chromosome 1q25-31; osteoarthritis; joint lubrication; osteopath;c;
KW antlarthritic.
XX
OS Homo sapiens.
XX
PN WC200107068-AI.
XX
PD 01-FEB-2001.
XX
PF 21-JUL-2000; 2000MC-US20002.
XX
PR 23-JUL-1999; 99US-0145328.
PR 19-JUL-2000; 2000US-0145328.
XX
PA (UYCA-) UNIV CASE WESTERN RESERVE.
XX
PI Warman ML;
XX
DR WPI: 2001-182721/18.
XX
PT New composition comprising the campodactylly-arthropathy-coxa
PT vara-pericarditis protein in combination with an anesthetic, useful for
PT treating osteoarthritis, or as lubricants of tissue and joints

DR N-PSDB: AAH98981.
 XX Isolated polypeptide for treatment of diseases, diagnostics, raising
 PT antibodies and research use -
 PT
 XX
 PS Claim 20; Page 1198-1201; 1275pp; English.
 XX
 CC The present invention provides the protein and coding sequences of novel
 CC proteins from a variety of organisms, including human, dog, cat, horse,
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
 CC from the organism of interest. They can be used in diagnostics,
 CC forensics, gene mapping, identification of mutations, to assess
 CC biodiversity and for nutritional purposes. The present sequence is a
 CC protein of the invention.
 CC
 XX
 XX Sequence 1299 AA:
 Query Match 100.0%; Score 5416; DB 22; Length 1299;
 Best Local Similarity 100.0%; Pred. No. 2.4e-298;
 Matches 1013; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VKDKKKNNRTKKKPKPPVNDAGSGLDNGDFKTTTDDTSTQTNKVSTSKITTAKPIN 60
 DB 200 vkdkkknnrtkkkpkppvndagsldngdfkvtltpdstlqtnkvsapklitakpin 259
 QY 61 PRPSLPNSDTSKETSJLVNKEIVETKETTNNKOTSTDGKEKTSKAKTOSIEKTSK 120
 DB 260 prpslpsndtsketsjlvnketivettntnkotstdgkettstaketsaketsak 319
 QY 121 DLATSVLVKKPPPKATETTTGKALITPKKEPTTPPKKPASTTPKEPTPTTISAPTPK 180
 DB 320 dlatstvlvkkpppkatetttgkaltltpkkepttpkkpastytpkeptpttisaptpk 379
 QY 181 EPAPTTKSAPTTKEBPATTTKEBPATTTKEBPATTTKBPATTTKSAPTTEKBPAPT 240
 DB 380 epapttksaptpkbpaptpkbpaptpkbpaptpkbpaptpkbpaptpkbpaptpk 439
 QY 241 PKRPAPTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKBPATTTKEBPAPT 300
 DB 440 pkraptpkbpaptpkbpaptpkbpaptpkbpaptpkbpaptpkbpaptpk 499
 QY 301 PTTPKKEBPATTTKEBPATTTKEBPATTTKSAPTTTKEBPATTTKSAPTTKEBPAPT 360
 DB 500 pttpkkebpaptpkbpaptpkbpaptpkbpaptpkbpaptpkbpaptpk 559
 QY 361 PAPTTPKEBPATTTKBPATTTKEBPATTTKEBPATTTKBPATTTKEBPAPT 420
 DB 560 papttpkebpaptpkbpaptpkbpaptpkbpaptpkbpaptpkbpaptpk 619
 QY 421 TPKKLTPTTPEKLAPTTPEKAPATTPEELAPTTPEEPTPTTPEEAPATTPKAAAPTPE 480
 DB 620 tpkkltpptpekapttpekapttpekapttpekapttpekapttpekapttpek 679
 QY 481 PAPTTPKEBPATTTKBPATTTKEBPATTTKBPATTTKBPATTTKBPATTTKEBPAPT 540
 DB 680 papttpkebpaptpkbpaptpkbpaptpkbpaptpkbpaptpkbpaptpk 739
 QY 541 EPSTTSDKRAPATTGKAPATTKEBPATTTKEBPATTTKBPATTTKEBPAPT 600
 DB 740 epsttsdkrapattgkapattkebpaptpkbpaptpkbpaptpkbpaptpk 799
 QY 601 KELAPTTTGGPTSTSDKAPATTPKETAPTTKEBPATTTKBPATTTPEPTTSEVST 660
 DB 800 kelaptttggptstsdkapattpketapttkebpaptpkbpaptpkbpaptpk 859
 QY 661 PTTTKEPTTIHKSDESTPELSAEPKALENSKEPGVPTTKTPATPEMTTAKKDT 720
 DB 860 ptttkepttihsdesptelsaepkalenskepgvpttktpatpemttaakkdt 919
 QY 721 TERDLATTEPTTAAKMTKETATTEKTESKITATTTQVSTTTQDTTPKITTLLKTT 780
 DB 721 terdlattepttAAKMTKETATTEKTESKITATTTQVSTTTQDTTPKITTLLKTT 780

DB 920 terdlrttpepttaapkmketalttektesklalttqvtstltdqtdtcfkittlktc 979
 QY 781 TLAPKVTTTKTTTITTEIMKPEETAKPKDRTATSKATTPKPKPTAKPKPTSTKPKPT 840
 DB 980 tlapkvtttktttiteimkpeetapkdkdrtatskattpkpkptakpkptstkkpkpt 1039
 QY 841 MPVRKRTKTPTPPKMTSTPELNPISRIAEAMLQTTTRPNQTPNSKLVEVNPSEDPAG 900
 DB 1040 mpvrkrtktptpkmtstpeelnpsriaeamlqtttrpnqtpnslvevnpkedag 1099
 QY 901 AEGTTPMLLRPHVEMPEVLPDDMDYLPRVNOGIIITPMISDEFNINCGKPVGGLTLRN 960
 DB 1100 aegttpmlrrphvempevlpddmdylprvnoigiitpmisdefnincgkpvvggltlrn 1159
 QY 961 GTLVAFRGHYFWMLSPPSPSPARRITVEWGISPIDVTFRNCBCKTPEFK 1013
 DB 1160 gtlvafrghyfwmllsppsparritvewgispidvtfrncbcktpfk 1212

RESULT 2
 AAR26049
 ID AAR26049 standard; Protein; 1404 AA.
 XX
 AC AAR26049;
 XX
 DT 02-FEB-1993 (first entry)
 XX
 DE MSF precursor.
 XX
 KW Megakaryocyte colony stimulating factor; secretion signal; meg-CSF;
 XX stability; proteolytic cleavage; adhesion; alternative splicing.
 OS Synthetic.
 OS
 FH Key
 FT Location/Qualifiers.
 FT 1..26
 FT /label= Exon_I
 FT 26..67
 FT /label= Exon_II
 FT 67..107
 FT /label= Exon_III
 FT 107..157
 FT /label= Exon_IV
 FT 157..200
 FT /label= Exon_V
 FT 200..1141
 FT /label= Exon_VI
 FT 1411..1166
 FT /label= Exon_VII
 FT 1166..1212
 FT /label= Exon_VIII
 FT 1213..1266
 FT /label= Exon_IX
 FT 1266..1331
 FT /label= Exon_X
 FT 1331..1373
 FT /label= Exon_XI
 FT 1373..1404
 FT /label= Exon_XII
 FT
 FT
 PN W09213075-A.
 XX
 PD 06-AUG-1992.
 XX
 PF 17-JAN-1992; 92WO-US00433.
 XX
 PR 18-JAN-1991; 91US-0643502.
 PR 10-SEP-1991; 91US-0757022.
 XX
 PA (GEMV) GENETICS INST INC.
 XX
 PI Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K;
 XX

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:16:33 ; Search time 107.17 Seconds
(without alignments)
700.161 Million cell updates/sec

Title: US-09-556-246-1_COPY_200_1212
Perfect score: 5416
Sequence: 1 VKDNKKRKRKKRPKPPVY.....SPIDVFRNCGKTFPRK 1013

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Genecod_1101.*
1: /SID8/gcgdata/geneseq/geneseq/AA1980.DAT:*
2: /SID8/gcgdata/geneseq/geneseq/AA1981.DAT:*
3: /SID8/gcgdata/geneseq/geneseq/AA1982.DAT:*
4: /SID8/gcgdata/geneseq/geneseq/AA1983.DAT:*
5: /SID8/gcgdata/geneseq/geneseq/AA1984.DAT:*
6: /SID8/gcgdata/geneseq/geneseq/AA1985.DAT:*
7: /SID8/gcgdata/geneseq/geneseq/AA1986.DAT:*
8: /SID8/gcgdata/geneseq/geneseq/AA1987.DAT:*
9: /SID8/gcgdata/geneseq/geneseq/AA1988.DAT:*
10: /SID8/gcgdata/geneseq/geneseq/AA1989.DAT:*
11: /SID8/gcgdata/geneseq/geneseq/AA1990.DAT:*
12: /SID8/gcgdata/geneseq/geneseq/AA1991.DAT:*
13: /SID8/gcgdata/geneseq/geneseq/AA1992.DAT:*
14: /SID8/gcgdata/geneseq/geneseq/AA1993.DAT:*
15: /SID8/gcgdata/geneseq/geneseq/AA1994.DAT:*
16: /SID8/gcgdata/geneseq/geneseq/AA1995.DAT:*
17: /SID8/gcgdata/geneseq/geneseq/AA1996.DAT:*
18: /SID8/gcgdata/geneseq/geneseq/AA1997.DAT:*
19: /SID8/gcgdata/geneseq/geneseq/AA1998.DAT:*
20: /SID8/gcgdata/geneseq/geneseq/AA1999.DAT:*
21: /SID8/gcgdata/geneseq/geneseq/AA2000.DAT:*
22: /SID8/gcgdata/geneseq/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5416	100.0	1299	22	AA24322 Human EST encoded
2	5416	100.0	1404	13	AA26049 MSF precursor.. Sy
3	5416	100.0	1404	22	AA260568 Human megakaryocyt
4	5416	100.0	1404	22	AA29773 Human megakaryocyt
5	3217.5	59.4	902	22	AA29778 Human MSF-derived
6	1357.5	25.1	452	16	AA24516 C899P predicted am
7	1168	21.6	5179	22	AA24516 Arabidopsis thalia
8	981	18.1	763	21	AA238942 C. thermocellum Ol
9	942	17.4	1664	19	AA43106 Bovine MSF ortholo
10	823.5	15.2	472	22	AA260569 Peptide #2327 enco
11	771.5	14.2	1325	22	AA260569 Peptide #2327 enco

12	715	13.2	763	18	AA231852 Mycobacterium tube
13	625.5	11.5	4412	21	AA233666 Sequence g1/101742
14	612	11.3	572	18	AA231855 Mycobacterium tube
15	544	10.0	844	7	AA260570 Sequence of the Pa
16	542	10.0	807	21	AA254467 Amino acid sequenc
17	520.5	9.6	788	21	AA254466 Cryptosporidium pa
18	508	9.4	1837	21	AA211726 Bioadhesive precu
19	506.5	9.4	744	9	AA282975 Human ORFX ORF995
20	486.5	9.2	2971	21	AA241231 Human SRCAP. Homo
21	489.5	9.0	2972	22	AA250362 Portion of CRYPTOS
22	489.5	9.0	3118	22	AA250362 P. yoelii SSP2 ant
23	489.5	9.0	1721	21	AA211727 Peptide #2892 enco
24	488.5	9.0	826	13	AA26042 Peptide #2869 enco
25	488	9.0	617	22	AA24187 Human MUC11 polype
26	488	9.0	957	21	AA259288 C900P predicted am
27	488	9.0	957	21	AA259288 Cryptosporidium pa
28	488	9.0	957	22	AA24513 Human protein sequ
29	488	9.0	1721	19	AA248299 Bioadhesive precu
30	485	8.8	1127	22	AA25541 Peptide #1317 enco
31	476.5	8.8	652	9	AA282974 Peptide #1289 enco
32	467.5	8.6	511	22	AA27312 Peptide #1289 enco
33	467.5	8.6	511	22	AA27312 Peptide #1289 enco
34	467.5	8.6	511	22	AA27312 Peptide #1289 enco
35	450.5	8.3	378	12	AA24160 Peptide #1289 enco
36	446.5	8.2	378	12	AA24160 Peptide #1289 enco
37	446.5	8.2	750	20	AA25477 Peptide #1289 enco
38	443	8.2	2870	21	AA25559 Peptide #1289 enco
39	443	8.2	3178	21	AA25559 Peptide #1289 enco
40	442.5	8.2	751	16	AA280839 Peptide #967 enco
41	433	8.0	2819	22	AA235408 Peptide #967 enco
42	428.5	7.9	2665	22	AA24533 Peptide #941 enco
43	428.5	7.9	2665	22	AA26950 Peptide #941 enco
44	428.5	7.9	2665	22	AA26950 Peptide #941 enco
45	428.5	7.9	3266	21	AA242491 Human ORFX ORF2255

ALIGNMENTS

RESULT 1	AA24322 standard: Protein; 1299 AA.
ID	AA24322
XX	AA24322
AC	AA24322
XX	AA24322
DE	12-OCT-2001 (first entry)
XX	Human EST encoded protein SEQ ID NO: 1847.
XX	Human: sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW	tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW	diagnostics; forensic test; gene mapping; genetic disorder;
KW	biodiversity; gene therapy; nutrition.
XX	
OS	Homo sapiens.
XX	
PN	WO200154477-A2.
PD	02-AUG-2001.
XX	
PF	25-JAN-2001; 2001WO-US02687.
XX	
PR	25-JAN-2000; 2000US-0491404.
PR	17-JUL-2000; 2000US-0617746.
PR	03-AUG-2000; 2000US-0631451.
PR	15-SEP-2000; 2000US-0663870.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI	Cao Y, Drmanac RA, Zhang J, Werhman T;
XX	
DR	WPI; 2001-476164/51.

Mon Apr 29 08:35:38 2002

us-09-556-246-1_copy_200_1263.rag

2 - 3

Page 16


```

OY 306 EP--APTTTKEPSTTKEPAPTTTTSAPTTTKEPAPTTTTSAPTTTKEPAPTTTKEPAP 363
Db 149 fppfpapalppap-----paplaanspplppapptpgrt--ppaapwpyvpaapkskspa 201
OY 364 TTPKEPAPTTTKEKAPPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 419
Db 202 spvpvppap-----pmatpmeifpplpvpvpdpisketppapppapppapppapppappp 256
OY 420 TTPKLTPTTPEKLA-----TTPKEPAPTTPEELAPTTPEEPPTTTEEPAPTT-- 469
Db 257 pvpnlipppap--apvavaavlvapcpplpplpnhppapppapvpyvplaplpnshp 313
OY 470 ---PKAAPNTPKKEPAP----- 483
Db 314 pappapavpyvplapplsipvsvwksftllstfcrcvsgvlagalapsrsapl 373
OY 484 ---TTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 533
Db 374 tttppalpaplpplpplpplntavpplpplpvtalaplplaplpplpplpplpplpplp 431
OY 534 LAPTTTKEPSTTSDKAPPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 587
Db 432 --pippgkpwltcpplapapepk-lvplpypgscppsekpnppapppapppapppappp 488
OY 588 KEAPAPTTT---KKAPKELAPTTTGP--TSTTSKAPAPTTTKEPAPTTTKEPAPTTTKEPAP 640
Db 489 appapmpsaavrvppspplpplpplpplpplpplpplpplpplpplpplpplpplppl 548
OY 641 KKAPAPTTTPTPTTSEVSTPTTKEPPTTIHKS-----DESPPELSAEPPTKA--LENSP 694
Db 549 --pappapptpklis--anppcpvypapnrvppapppapppapppapppapppapppap 604
OY 695 KEGVPTTPTATKPKEMTTTAKDKTTERDLRTPETTTAPKMKETATTEKTESKI 754
Db 605 --pappapppapalpfvppa-----ppupaak-----srpal 637
OY 755 TATTTTGVSTTTQDTPPKITTLKTTLAPKVTYTTTKTTTTEIMKPKETAKPKMRATN 814
Db 638 paappapppavrvatp-----ppappapppap 665
OY 815 SKATTPKPKP-----TKAPKPTSTKPKTPMVRKPKTTTPTRKMTSTM-----PELNP 865
Db 666 smaipapppdpplpplpplpplpplpplpplpplpplpplpplpplpplpplpplppl 725
OY 866 TSRIAEAMLQTTTRPNQTPNSKLIVENPKSEDAEGAEGET 905
Db 726 vppaplaplpingrvfarknsl-----gsssgdt 756

```

```

RESULT 13
AA53666
ID AA53666 standard; Protein: 4412 AA.
XX
AC AA53666;
XX
DT 22-FEB-2000 (first entry)
XX
DE Sequence gi/1017427/emb/CAA62189 from an alignment with protein 608.
XX
KM Mechanical stress; gene therapy; protein 608; osteoporosis; bone density;
XX
KW bone development; gi/1017427/emb/CAA62189.
XX
OS Unidentified.
XX
PN W0960164-A1.
XX
PD 25-NOV-1999.
XX
PF 14-MAY-1999; 99MO-US11066.
XX
PR 15-MAY-1998; 98US-0085673.
XX

```

```

PA (ODAR-) QUARK BIOTECH INC.
XX
PI Elnat P, Mor O, Skallier R, Feinstein E, Faerman A;
XX
DR WPI; 2000-053304/04.
XX
PT Identification of stress induced genes for determining risk and
XX
PS preventing, treating or controlling osteoporosis
XX
PS Claim 32; Fig 6A-R; 308pp; English.
XX
CC The present sequence is obtained from a clustal X alignment with
CC protein 608. Protein 608 was identified using the method of the invention
CC after subjecting rat osteoblasts to mechanical stress. Expression of the
CC 608 gene was found to be upregulated by about 3-fold in cells subjected
CC to mechanical strain. The specification describes a method for the
CC identification of genes responsive to a specific mechanical stress. The
CC method comprises applying the mechanical stress to an organism (tissue
CC or cells comprising bone cells), isolating the specific cellular
CC fractions and extracting mRNA from them, and differentially analysing the
CC mRNA in comparison with control samples. The method is used to identify
CC genes whose expression is responsive to a specific stress. The identified
CC genes are employed in determining risk associated with a physiological or
CC disease state. The risk determination methods are used for testing a
CC medicament for gene therapy. These medicaments, or genes identified by
CC the method of the invention, are used for treating, preventing or
CC controlling a physiological or disease state (especially osteoporosis or
CC bone density) or other factors causing or contributing to osteoporosis or
CC its symptoms or other conditions involved in mechanical stress or its
CC lack. The methods can also be used for advancing research or studies in
CC bone development.
XX
SO Sequence 4412 AA;

```

```

Query Match 11.0%; Score 625.5; DB 21; Length 4412;
Best Local Similarity 26.2%; Pred. No. 2.5e-27;
Matches 225; Conservative 86; Mismatches 379; Indels 169; Gaps 40;

```

```

OY 99 TDGKKEKTSKKESTQSIETK-----SAKDLAPTSTKVLAKPPKAETTTKGPALN----- 146
Db 3515 tdfgrllgrveehvvekvhyrvfeaevevfkpapp-----kgeisekllp 3566
OY 147 TPKKEPPTTP--KEPASTPKKEPTPTTTSAPTTKEP--APTTTTSAPTTTKEPAPTTTKE 204
Db 3567 pkkpkkvvpkkvvpkkvvpkkvvpkkvvpkkvvpkkvvpkkvvpkkvvpkkvvpkkvvp 3621
OY 205 PAPTTPKEPAPTTTKEPAPTTTTSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 260
Db 3622 ---ekkvvpvppakvppap-----ppkvpeapkevvpekkvvpvppkkvpevptkvpev 3672
OY 261 TPTTKEPAPTTTKEPAPTTT-----KEPAPTAEPKAPPTTKEPAPTTTKEPAPTTTKE 313
Db 3673 kaavpkkvpeaipkpepppvefeepsepsapkk--kpevpvrvpvpkvevpekkv 3731
OY 314 EPSPTTPKEPAPTTTKEPAPTTTKEPAPTTTTSAPTTTKEPAPTTTKEPAPTTTKEPAPTT 373
Db 3732 paap--pkkpvevrvk--vpeapkevvpekkvvpv--pkkkvevvpkt-----vpevkvav 3782
OY 374 PKPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 429
Db 3783 pekkvp-----ealppkpespppvefeepsealeppaevveeppaapvpyvtpknpv 3838
OY 430 PEKLAPTT-----PEKAPPTTPEELAPTTPEEPPT--TTPKEPAPTTTKEPAPTTTKEPAP 483
Db 3839 pekkapavvakkpelpvkvpevpekkvplvvpkk--peapkvavpevpekkv 3896
OY 484 ---TTPKEP-----APTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 534
Db 3897 kkvavpkkpvpvppavvvpkkv-----leekpavp-----vpeaesppvevvee--peel 3947
OY 535 APTTTKEPSTTSDKAPPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 594

```

```

Db 60 nctrhskpdkptgnsktldhk-saldhheappteensngqkdpmlrngsvdpsadt 118
Qy 56 A-----KPINPRPSLPPN---SDTSKET-----SLVUNKETVETKET 90
Db 119 tthkesagkhiltpapkskincrstskstgskvtcrskdctgrpleksmslaktstskt 178
Qy 91 TTT-----NKOTSDGKEKTSKAKET-----OSIEKT-SAKDLAPTSKVLAKPT 133
Db 179 ttfsmngsqckqskstafpekltlaasktyktgtgpeesektedstvtssdkllkctt 238
Qy 134 PKAETTTGCPALT-----TPKEPTPT---TPKEPASTTKEPTPT 170
Db 239 knlgecllsaneltqslaeptehegrrtanennlpsaeptenrerlanentlpsaepten 298
Qy 171 TTKSA--PTTPKEBAPT-----TTKSAPTTPKEBAPTTPKE---PATTPKEBAPT 216
Db 299 rerlanentapfapgtcenremtanentllfpaeppteigertanentlpsaepteheg 358
Qy 217 TTKEBAPTTPKSAPTTPKEBAPTTPKKAPATTTPKEBAPTTPKEPTP---TPKEBAPT 271
Db 359 tane-----ntlpsaeptehegertfandktssaeptehegertplanentlpsaept 414
Qy 272 -TKBAPTTPKEBAPTAPKKAPATTTPKEBAPTTPKEPA---PTTTKEBAPTTPKEPA--- 324
Db 415 enertanentlpsaeptenrerlanentlpsaeptenrerlanentlpsaepteng 474
Qy 325 -----PTTTKSA-PTTTKEBAPTTPKSAPTTPKEBAPTTPKE-----PATTPK 367
Db 475 qtrpfanektssaeptehegertplanen--tlpsaeptenrerlanentlpsaept 532
Qy 368 EPAPTTPKKAPATTTPKEBAPT-----TPKEBAPTTPKKAPATAPKEBAPTTPKEATTP 422
Db 533 engdrtpplanektlpslaeptengqtrpfanektssaeptehee---rtplanentlp 589
Qy 423 KLLPTTPPEKLAPTTPKEBAPTTPPEELAPTPPEEPTTPPEBAPTTPKAAAPT---PK 479
Db 590 spaepcenrerlanentlpsaeptenremtanentllfpaepcenrerlanektsspa 649
Qy 480 EPAPTTPKEBAPTTPKEBAPTTPKEBAPTTPPKGAPTTLKEBAP-----TPPKAPAKE 533
Db 650 eptengqtrpfanektssaeptehegertplanentlpsaeptenrerlanektppf 709
Qy 534 LAPTTKEPTSTSDKAPATTTPKGAPTTPKEBAPTTPKEBAPT-----TPKGAPTTLK 588
Db 710 aeptenrerlanentlpsaeptengqtrpfanektlpslaeptengqtrpfanektss 769
Qy 589 EPAPT-----TPKKPA-PKE-----LAPTTKGPTSTSDKAP 621
Db 770 saepptehegertplanentlpsaeptenrerlanentlqfpaepptehegertlanektppf 829
Qy 622 TTPKET-----APTTPKEBAPTTPKKAPATTTPETPTTSEVSTPTTKEPTTI 670
Db 830 aeptenrerlanentlpsaeptehegertplanentlpsaeptehegertfnektlps 889
Qy 671 HKSPDE-----STPELSAEPKPK---ALENSPKPEGVPT-----TKTPA 706
Db 890 saepptehegertplanentlps--raepptehegertlanekapcpkcphegertlps 948
Qy 707 ATP-----EMTTAKDKTTERDLRTTPETTTAPKMKKE---TATTEKTE 751
Db 949 saepptehegertplanentlpsaeptehegertlanentlpsaeptehegertlanektlp 1008
Qy 752 SKTTATT-TOVSTTODTTPFKI-----TTL-----KTTTLAPK 785
Db 1009 spaepptehegertlanentlpsaeptehegertlanentlpsaeptehegertlanekt 1068
Qy 786 VTTT-----KTTTTEIMNKPEETAKPKDRATNSKATTPKQKP----- 825
Db 1069 tlpslakptehegertlpspdktssaeptehegertlanentlpsaeptehegertlanekt 1128
Qy 826 -----TKAPKPTSTSKKPKTPRVKRP-----KTTTPPKWTSTMPKL 863
Db 1129 ekmtgvtkstehpkektstte-ktlrtpckptlysekctickgknlcppekptenlgn 1187

```

```

Qy 864 NPTSRIAEMIQTTTRPNOT 883
Db 1188 tlletlkapvkslenpekt 1207

RESULT 12
AAW31852 standard; Protein; 763 AA.
AAW31852:
27-APR-1998 (first entry)
DE Mycobacterium tuberculosis 74 kDa protein.
KW Tuberculosis; mycobacteria; infection; diagnosis;
KM antimycobacterial; antibiotic; vaccine.
XX Mycobacterium tuberculosis.
OS
XX
XX WO9741252-A2.
XX
XX 06-NOV-1997.
XX
XX 18-APR-1997; 97WO-EP01973.
XX
XX 29-APR-1996; 96DE-4017184.
XX
XX (GBFB ) GBF GES BIOTECH FORSCHUNG GMBH.
XX
XX Espilitia C, Honisch C, Moreno C, Singh M;
XX
XX WPI; 1997-549750/50.
XX
XX N-PSDB; AAT93610.
XX
XX New DNA and related proteins or RNA derived from M. tuberculosis -
XX used for diagnosis of mycobacterial infections, monitoring
XX vaccination and development of anti-mycobacterial agents
XX
XX Claim 5; Fig 13; 55pp; English.
XX
XX This novel 74 kDa protein is encoded by an open reading frame of
XX a Mycobacterium tuberculosis DNA fragment (see AAT93610) containing
XX a polymorphic GC-rich sequence. Its amino acid sequence shows
XX a high proline content, but there is no homology to any known
XX proline-rich antigens of mycobacteria. Novel M. tuberculosis
XX proteins (see AAW31851-57) are claimed. These can be produced as
XX recombinant proteins, especially in bacterial, yeast, fungal or
XX higher eukaryote host cells, and used for diagnosing tuberculosis
XX and other mycobacterial infections in humans or animals. The
XX claimed proteins can also be used for epidemiological studies, for
XX monitoring vaccination, and for the development of vaccines and
XX anti-mycobacterial drugs.
XX
XX Sequence 763 AA:

Query Match 12.5%; Score 715; DB 18; Length 763;
Best Local Similarity 28.2%; Pred. No. 3,4e-33;
Matches 248; Conservative 39; Mismatches 361; Indels 232; Gaps 39;

Qy 132 PTPKAEETTTGCPALTTPKEBAPTTPTPKEBAPTTPKPEPTTTISAPTTPKEBAPTTPKSAP 191
Db 3 pvp-----apralaplpapppapapkskppfppap-----pappcmllvsap 46
Qy 192 TTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKSAPTTPKEBAPTTPKKAPATTTPKE 251
Db 47 pcp--pap-----papkpkskafpfpvpappparelappip--pap-----peapre 90
Qy 252 PAPTTTPKEPTPTT-----KPEBAPTTPKEBAPTTPK-EPAPTAPK-KPAPTTPKEBAPTTPK 305
Db 91 srpalpccppppvvlpdppeapapvpvapnspfpfpfpfpapktvpapvpv--pvpnsp 148

```



```

PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161350.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 17.2%; Score 981; DB 21; Length 763;
 Best Local Similarity 38.5%; Pred. No. 2.6e-48;
 Matches 226; Conservative 24; Mismatches 293; Indels 44; Gaps 11;

```

QY 137 ETTTGGPALTTPKKEPTTPPKKEPASTTPKKEPTTTTTSAPTTPKKEPAPTTTTSAPTTPKE 196
DB 126 qtdqgnrlylrppprlpppcvncp-esp-----pppvtc--pqlpvt 170
QY 197 PATTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTT--TKSAPTTPKKEPAPTTTPKKEPAPTT 254
DB 171 lptltlppclpplcpatpctevlptqpltpqtlppltpclppltpclppltpclppltpclp 226
QY 255 TTPKKEPTTPTPKKEPAPTT--KEPAPTTTPKKEPAPTPAKKRAPTTTPKKEPAPTTTPKKEPAPTT 313
DB 227 nuppcppltpclppltpclppltpclppltpclppltpclppltpclppltpclppltpclp 286
QY 314 EPSPTTPKKEPAPTTTTSAPTTPKKEPAPTTTTSAPTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTT 373
DB 287 etepnclppltpclppltpclppltpclppltpclppltpclppltpclppltpclppltpclp 340
QY 374 PKKRAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTT 433
DB 341 ppgtclppltpclppltpclppltpclppltpclppltpclppltpclppltpclppltpclp 400
QY 434 APPTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTT 493
DB 401 lntclppltpclppltpclppltpclppltpclppltpclppltpclppltpclppltpclp 460
QY 494 PKKRAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTT 553
DB 461 pkltp-cltpclppltpclppltpclppltpclppltpclppltpclppltpclppltpclp 514
QY 554 TPKGTAPTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTT 613
DB 515 tppgclppltpclppltpclppltpclppltpclppltpclppltpclppltpclppltpclp 574

```

```

QY 614 TTSDKP-----APTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTT 666
DB 575 lppntclppltpclppltpclppltpclppltpclppltpclppltpclppltpclppltpclp 634
QY 667 P---TTTHKSPDESTEPKSAEPTPKALENSPKKEGVTTPKKAATP 710
DB 635 pppcltppltpclppltpclppltpclppltpclppltpclppltpclppltpclppltpclp 681

RESULT 10
AAW43106
AAW43106 standard; Protein; 1664 AA.
AAW43106;
XX
AC AAW43106;
XX
XX 16-OCT-1998 (first entry)
DT
XX C. thermocellum O1pB protein.
DE
XX
XX Multimer; enzyme; complex; protein-protein interaction; dockerin domain;
KW cohesin domain; catalytic subunit; scaffold subunit; sdb; synergistic;
KW cellulosome integrating protein; scaffoldin dockerin binding protein.
XX
OS Clostridium thermocellum.
XX
FH Key location/Qualifiers
FT Domain 28..192
FT Domain /note="cohesin type II domain"
FT Domain 207..363
FT Domain /note="cohesin type II domain"
FT Domain 409..565
FT Domain /note="cohesin type II domain"
FT Domain 607..763
FT Domain /note="cohesin type II domain"
XX
XX FR2748479-A1.
XX
XX 14-NOV-1997.
XX
XX 10-MAY-1996; 96FR-0005854.
XX
XX 10-MAY-1996; 96FR-0005854.
XX
XX (INSP ) INST PASTEUR.
XX
XX Beguin P, Leibovitz E;
DR WPI; 1998-011569/02.
DR N-PSDB; AAT86623.
XX
XX Cellulase proteins with cohesin or dockerin type II domains - useful
XX for potentiating the activity of multiprotein enzyme complexes
XX
XX Claim 7; Page 31-39; 60pp; French.
XX
XX Multimeric protein, especially enzymatic, complexes are held together
XX by protein-protein interactions between domains designated dockerins
XX and cohesins, which are found on the catalytic and scaffold subunits
XX respectively. An example of such a complex is the cellulose degrading
XX protein complex from Clostridium thermocellum, known as the cellulosome.
XX This complex comprises around 15 proteins including endoglucanases,
XX cellobiohydrolases, hemicellulases, e.g. xylanases or lichenases, which
XX interact with a central "scaffold" protein designated the cellulosome
XX integrating protein (CIPa; see AAW43108). The catalytic subunits
XX interact with the CIP subunit via conserved 23 amino acid dockerin
XX domains. CIP has been shown to contain 9 copies of a cohesin domain.
XX The invention relates to the isolation of proteins binding to a novel
XX dockerin type domain found in the C-terminal portion of CIP. The new
XX domain is designated a type II dockerin domain (as compared to the type
XX I domain found on the catalytic subunits of the cellulosome). The type
XX II dockerin domain has some sequence similarity to the type I dockerins
XX but is unable to bind type I cohesin domains.
XX The sequence presented here is an example of a protein which binds

```

PF 25-FEB-2000; 2000EP-0301439.
 XX 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123348.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130447.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 04-MAY-1999; 99US-0132407.
 PR 05-MAY-1999; 99US-0132484.
 PR 06-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 06-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 14-MAY-1999; 99US-0134768.
 PR 18-MAY-1999; 99US-0134941.
 PR 19-MAY-1999; 99US-0135124.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140333.
 PR 23-JUN-1999; 99US-0140334.
 PR 24-JUN-1999; 99US-0140655.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.

PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143342.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144864.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 03-AUG-1999; 99US-0146389.
 PR 04-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 05-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151338.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 11-SEP-1999; 99US-0153758.
 PR 13-SEP-1999; 99US-0154018.
 PR 15-SEP-1999; 99US-0154039.
 PR 16-SEP-1999; 99US-0154779.
 PR 20-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.


```

Db 1770 fspfsttpttbpvclcmwlgwldsgkpnfhkpggdcltlgdvcgpgwaaniscratmvp 1829
Qy 428 -----TTPKLAAPT 437
Db 1830 dvpiglgqgvcdvsyglcknedqkpgvlpmafclynelnvqcecvlqptlmtlctt 1889
Qy 438 PEKPAPTTPEELAPTTPEEPPTT-TPPEAPATTTPKAAAPTTPKEAPATTKEP-----AP 491
Db 1890 temppttpttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1949
Qy 492 TTPKEAPATT-TPKEAPATTTPKGAATTLKAPATTTPKAPKELAPTTTKEPTSTSDK 550
Db 1950 tctvpttpttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 2009
Qy 551 AP-TTPKGAATTPKAPATTTPKGAATTLKAPATTTPKAPKELAPTTT 609
Db 2010 tpttpttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 2068
Qy 610 GPTSTSTDAKAP-TTPKEAPATTTPKAPATTTPKAPATTTPPTTSEVSTPTTKEPT 668
Db 2069 tpttpttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 2127
Qy 669 TIKSPDESTPELSAETPKALENSPKP-----GVPT-TTPPATKPEMTTAK 717
Db 2128 t-----tpttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 2184
Qy 718 DKTTENDLR---TTP--ETTTAAPKMT---KETATTTETKTESKITATTTQVSTTTODT 769
Db 2185 ptpcttpttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 2243
Qy 770 TPKKITTLKTTTLAPVTTT-KTITTTTELMMKPEETAKPKDAINASKATTPKPKQPTKA 828
Db 2244 tp--tcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 2299
Qy 829 P-KKPTSTKKPKMTTPVRRKPTTTPTRKMTSTWPELNP---TSRIAEAMLOTTTR-PNQT 883
Db 2300 ptpcttpttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 2359
Qy 884 PMSKLVAVPKSEDAAGCEPTPHMLLRPHVEMPEVTP 921
Db 2360 ptttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 2392

RESULT 8
AAB60569
ID AAB60569 standard; Protein; 472 AA.
AC AAB60569;
XX
XX
XX 27-APR-2001 (first entry)
XX
XX
XX Bovine MSF orthologue, superficial zone protein (SZP).
XX
XX
XX Bovine; CACP protein; camptodactylly-arthropathy-coxa vara-pericarditis;
XX
XX
XX superficial zone protein; SZP; MSF orthologue; synovial lubricant;
XX
XX
XX osteoarthritis; joint lubrication; osteopathic; antiarthritic.
XX
XX
XX Bos taurus.
XX
XX
XX WO200107068-A1.
XX
XX
XX 01-FEB-2001.
XX
XX
XX 21-JUL-2000; 2000WO-US20002.
XX
XX
XX 23-JUL-1999; 98US-0145328.
XX
XX
XX 19-JUL-2000; 2000US-0145328.
XX
XX
XX (UYCA-) UNIV CASE WESTERN RESERVE.
XX
XX
XX Warman ML;
XX
XX
XX WPI: 2001-182721/18.

```

XX New composition comprising the camptodactylly-arthropathy-coxa
PT vara-pericarditis protein in combination with an anesthetic, useful for
PT treating osteoarthritis, or as lubricants of tissue and joints
XX
XX Example 1; Fig 4; 34pp; English.

CC The invention relates to a method of treating osteoarthritis via the
CC administration of a composition comprising the camptodactylly-arthropathy-
CC coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.
CC The composition may further comprise a local anesthetic. The composition
CC of the invention may be administered via intra-articular or intravenous
CC injection. The human CACP protein is identified in the invention as
CC being megakaryocyte stimulating factor (MSF). The gene encoding
CC CACP protein (MSF) is located on chromosome 1q25-31, and mutations in
CC this gene are responsible for the heritable disorder camptodactylly-
CC arthropathy-coxa vara-pericarditis, in which patients have synovial
CC hyperplasia without evidence of inflammation. CACP protein (MSF)
CC acts as a synovium lubricant, and can be used to lubricate tissue and
CC joints in the treatment of osteoarthritis. The composition may be
CC applied to reduce the symptoms of osteoarthritis (e.g., joint pain,
CC loss of range of movement or joint damage). The present sequence
CC represents the bovine orthologue of human MSF, superficial zone
CC protein (SZP).
XX

Sequence 472 AA:

Query Match 18.3%; Score 1042.5; DB 22; Length 472;
Best Local Similarity 76.4%; Pred. NO. 5e-52;
Matches 197; Conservative 20; Mismatches 38; Indels 3; Gaps 3;

```

Qy 808 PDORATNSKATTPKOKKPKAKKPTSTKKTMRVRRKPTTPPRK-MSTWPELNP 866
Db 76 pgratnsqvtltpkqkpkpkpkstktprt-privtkpttcttpttcttcttctt 134
Qy 867 SRIAEAMLOTTTRPQNTSPKLEVNPKSEDAAGCEPTPHMLLRPHVEMPEVTPDMOYL 926
Db 135 s-lpeamqlttrpttpttpttpttpttpttpttpttpttpttpttpttptt 193
Qy 927 PRVPTQGIITNMLDENICNGKFPVDGLTTLRNGTLVAFRGHYFWMLSPPSPSPARRI 986
Db 194 vtrpsqgfglpmfisdetnlcmgrpvdgltlrlngllvalfghyfwmlptftpppprrl 253
Qy 987 TEVWGIPSPIDVFTFRNCEGKTFEFGKSOYRFTNDIKDGYPRIRKGGGLTGQIYA 1046
Db 254 tevwgipspidvftfrnccegtftfkgsgywrftndikdagypkllskfgglngkiva 313
Qy 1047 AISTAKYKNWPESEVYFEK 1064
Db 314 alslagyskrspevyffk 331

RESULT 9
AAG38942
ID AAG38942 standard; Protein; 763 AA.
AC AAG38942;
XX
XX
XX 18-OCT-2000 (first entry)
XX
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 48115.
XX
XX
XX Protein identification; signal transduction pathway; metabolic pathway;  

XX
XX
XX hybridisation assay; genetic mapping; gene expression control; promoter;  

XX
XX
XX Arabidopsis thaliana.
XX
XX
XX EPI033405-A2.
XX
XX
XX 06-SEP-2000.
XX
XX
XX

```

CC shown in AAT04544 and AAT04545. The fragments amplified by these primers
CC can then be used as probes to screen human cDNA libraries for MPO cDNA.
CC The MPO cDNA can then be inserted into a plasmid which is used to
CC transform cells to produce MO. The MO sequence is capable of promoting
CC colony formation of megakaryocytes, enlarging the size of megakaryocytes
CC and stimulating the proliferation of multipotential stem cells. The
CC factor may be used for treating thrombocytopenia and hematocytopenia.
CC The purification method can be used to isolate MPO from human urine or
CC serum of patients with aplastic anaemia, and from animal blood or urine
CC by radiation exposing the animals to induce aplastic anaemia.

XX Sequence 452 AA;

Query Match 23.8%; Score 1357.5; DB 16; Length 452;
Best Local Similarity 88.3%; Pred. No. 5.9e-70;
Matches 263; Conservative 3; Mismatches 21; Indels 11; Gaps 1;

QY 1 VKDNKKNTTKKKPPKPPVVDAGSGLDNGDFKVTPTSTTQNHKKVTSKRTAKPIN 60
DB 157 vkdnknttkkkppkppvvdagsgldngdfkvtptsttqnhkvtspkltakpIn 216
QY 61 PRPSLPNSDSKESLTVNKETTYETKETTNNKQSTDGKEKTSKETSIEKTSK 120
DB 217 prpslpnsdskseltvnkettvetkettntkqsttdgketsketsktsak 276
QY 121 DLAPTSKVLAKPPTKPAETTTGAPALYTPKEPTTTPKPEASTTKEPTTTTAKSAPTPK 180
DB 277 dlaptskvlakpptaetttgpaltpkpttptkpeasttkpeptttksaptpk 336
QY 181 EPAPTTTSAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTT 240
DB 337 epaptttsapttkpeapttkpeapttkpeapttkpeapttkpeapttkshpjlpcscxxx 396
QY 241 EKKPAPPTPKPEAPPTPKPEPTTTPKPEAPTTKPEAPTTKPEAPTTKPE 298
DB 397 ctqp-----tpkehpplprslhpkpkeapptkpeaptpakpaplplle 443

RESULT 7
AAM24516
ID AAM24516 standard; Protein; 5179 AA.

XX AAM24516;
XX
XX 12-OCT-2001 (first entry)
XX
XX C899P predicted amino acid sequence.
XX
XX Human; immunotherapy; diagnosis; colon cancer; colon tumour;
XX immunogenic; gene therapy; vaccine; colonic cancer.
XX
XX Homo sapiens.
XX
XX WO200149716-A2.
XX
XX 12-JUL-2001.
XX
XX 29-DEC-2000; 2000WO-US35596.
XX
XX 30-DEC-1999; 99US-0476296.
XX 10-JAN-2000; 2000US-0480321.
XX 15-FEB-2000; 2000US-0504629.
XX 06-MAR-2000; 2000US-0519444.
XX 19-MAY-2000; 2000US-0575251.
XX 29-JUN-2000; 2000US-0609448.
XX 28-AUG-2000; 2000US-0649811.
XX
XX (CORI-) CORIXA CORP.
XX
XX Xu J, Lodes MJ, Secretist H, Benson DR, Meagher MJ, Stolk JA;
XX King GE, Wang T, Jiang Y;
XX

DR WPI; 2001-441847/47.

XX Colon tumor associated proteins and nucleic acids useful for the
PT prevention, diagnosis and treatment of colonic cancer

XX Claim 2; Page 446-462; 472pp; English.

XX The present invention describes colon tumor associated proteins (I) and
CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
CC (II) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate colon tumor associated protein (TCAP)
CC expression, such as colonic cancer. For example, (I) and (II) may be
CC used to treat disorders associated with decreased expression by
CC rectifying mutations or deletions in a patient's genome that affect the
CC activity of TCAPs by expressing inactive proteins or to supplement the
CC patients own production of them. Additionally, (II) may be used to
CC produce the TCAP proteins, by inserting the nucleic acids into a host
CC cell culturing the cell to express the protein. (II) and its
CC complementary sequences may also be used as DNA probes in diagnostic
CC polymerase chain reaction (PCR) and hybridisation assays to detect and
CC quantitate the presence of similar nucleic acids in samples, and
CC therefore which patients may be in need of restorative therapy. (I) may
CC also be used as antigens in the production of antibodies against TCAPs
CC and in assays to identify modulators of TCAP expression and activity.
CC Anti-(I) antibodies and antagonists may also be used to down regulate
CC TCAP expression and activity. The anti-(I) antibodies may also be used
CC as diagnostic agents for detecting the presence of TCAPs in samples
CC (e.g. by enzyme linked immunosorbent assay (ELISA)). AAI28460 to AAI29512
CC and AAM24494 to AAM24533 represent nucleotide and amino acid sequences
CC given in the exemplification of the present invention.

SO Sequence 5179 AA;

Query Match 20.5%; Score 1168; DB 22; Length 5179;
Best Local Similarity 32.5%; Pred. No. 4.1e-58;
Matches 363; Conservative 53; Mismatches 448; Indels 254; Gaps 40;

QY 21 DEAGSGLDNGDFK-----VTPTSTTQNH-KKVTSPK----- 52
DB 1312 dhpssgsdgdgdepfdygcgapediecrsvkqphslqhgqkvcdsvfgicknedgf 1371
QY 53 -----TTAKPINRPSLPNSDSKESLTVNKETTYETK 88
DB 1372 gngpfglcydykivrncwpmkcltptpspttlllpptpsptt----- 1427
QY 89 ETTTNNKQSTDGKEKTSKETSIEKTSKADLAPTSKVLAKPPTKPAETTTGAPALTP 148
DB 1428 -tttpppttsppttltt-----lpt-----tpspjstlttptt 1470
QY 149 KEPT-----PTPKPEASTTTPKPEPTTTPKPE-----APTTPKSAP-TTP 194
DB 1471 sppttsppttsppttlttpttpttsppttlttpttsppttlttpttspptt 1530
QY 195 KEAPPTTKPEAPTPKPE-----APTTPKEAPTTTKSAP--TTPKEAPTPPKRA 245
DB 1531 ttppttsppttsppttlttpttstltlpttsppttlttpttsppttsppt 1590
QY 246 PTPPKPEAPTTKPEPTTTPKPEAPTTKPEAPTTKPE-----APTAKKRAPTPKE 298
DB 1591 tttpppttsppttlttpttsppttsppttlttpttstltlpttsppttltt 1650
QY 299 PAPTTPKEAPTTTKEPSTTPKPEAPTTTTPKSAPTTTPKSAPTTTPKSA--TTPKEPSP 355
DB 1651 pppttsppttsppttlttpttsppttsppttsppttlttpttsppttsppt 1709
QY 356 TTPKPEAPTTKPEAPTPPKKAPPTTPKPEAPTTTPKKAAPTPKAPKAPATTP- 414
DB 1710 tttptstlttsppttlttpttsppttsppttlttpttsppttlttpttsppt 1769
QY 415 -----KETAPTP-----KKLTP 427

CC Lubricating moiety and has a polypeptide sequence comprising 1-76
CC repeats of a motif having at least 50% identity to the sequence KEPAPPT
CC (AAB293774). The invention also relates to a nucleic acid encoding a
CC human MSF-derived tribonectin; a biocompatible composition comprising a
CC human tribonectin for inhibiting tissue adhesion formation; and a method
CC of diagnosing osteoarthritis or a predisposition to osteoarthritis by
CC measuring the amount of MSF or its fragment in a biological sample of a
CC mammal, wherein an increased amount of MSF compared to a control
CC indicates the presence of or predisposition to developing
CC osteoarthritis. The tribonectin and DNA encoding it are useful in the
CC treatment of osteoarthritis, where they may be used for lubricating
CC mammalian joints, such as articulating joints of humans, dogs or horses.
CC The tribonectin, when formulated as a membrane, foam, gel or fibre, is
CC useful for inhibiting adhesion between two surfaces such as the injured
CC tissues of a mammal, where the injury is caused by a surgical incision
CC or trauma, or an artificial device e.g., an orthopaedic implant. In
CC particular, one of the surfaces is pericardial tissue. DNA encoding a
CC tribonectin may be used in gene therapy. The present sequence represents
CC a substantial portion of a human MSF-derived tribonectin.

SQ Sequence 902 AA;

Query Match	61.18	Score 3484	DB 22	Length 902
Best Local Similarity	74.68	Pred. No. 1.4e-190		
Matches 794	Conservative 28	Mismatches 80	Indels 162	Gaps 66

[illegible][illegible]

XX	RESULT	6
XX	AA80041	
XX	ID	AA80041 standard; Protein: 452 AA.
XX	AC	
XX	AA80041;	
XX	DT	10-APR-1996 (first entry)
XX	DE	Human megakaryocytopoietin protein.
XX	DE	
XX	KW	Human; megakaryocytopoietin; wheat germ agglutinin; heparin;
XX	KW	megakaryocyte; aplastic anaemia; thrombocytopenia; hematocytopenia;
XX	KW	multipotential stem cell.
XX	OS	Homo sapiens.
XX	Key	Location/Qualifiers
XX	FT	Misc-difference 393..396
XX	FT	/note="unspecified amino acids"
XX	FT	Misc-difference 444..446
XX	FT	/note="unspecified amino acids"
XX	PN	W09523861-A1.
XX	PD	08-SEP-1995.
XX	PF	06-MAR-1995; 95WO-CND0015.
XX	PR	04-MAR-1994; 94CN-0112066.
XX	PA	(SHAN-) SHANGHAI BEITI BIOTECHNOLOGY CO LTD.
XX	PI	Gu X, Han Z, Shen Q;
XX	DR	WPI: 1995-320576/41.
XX	DR	N-PSDB; AAT04546.
XX	PT	New haematopoietic cell growth factor - used for treating
XX	PT	thrombocytopenia and hematocytopenia
XX	PS	Example; Page 23; 36pp; Chinese.
CC	CC	This sequence represents the human megakaryocytopoietin (MPO) protein.
CC	CC	This sequence was purified using a carrier which can couple wheat germ
CC	CC	agglutinin and heparin to separate MPO. Fragments of this sequence (see
CC	CC	AA80039 and AA80040) were used to produce the amplification primers

XX Claim 3; Page 7; 47pp; English.

PS
XX The invention relates to a human tribonectin which is a product of
CC alternative splicing of the human MSF (megakaryocyte stimulating factor)
CC gene. The tribonectin has at least one O-linked oligosaccharide
CC indicating moiety and has a polypeptide sequence comprising 1-76
CC repeats of a motif having at least 50% identity to the sequence KRPAPT
CC (AAB29374). The invention also relates to a nucleic acid encoding a
CC human MSF-derived tribonectin; a biocompatible composition comprising a
CC human tribonectin for inhibiting tissue adhesion formation; and a method
CC of diagnosing osteoarthritis or a predisposition to osteoarthritis by
CC measuring the amount of MSF or its fragment in a biological sample of a
CC mammal, wherein an increased amount of MSF compared to a control
CC indicates the presence of or predisposition to developing
CC osteoarthritis. The tribonectin and DNA encoding it are useful in the
CC treatment of osteoarthritis, where they may be used for lubricating
CC mammalian joints, such as articulating joints of humans, dogs or horses.
CC The tribonectin, when formulated as a membrane, foam, gel or fibre, is
CC useful for inhibiting adhesion between two surfaces such as the injured
CC tissues of a mammal, where the injury is caused by a surgical insertion
CC or trauma, or an artificial device e.g., an orthopaedic implant. In
CC particular, one of the surfaces is pericardial tissue. DNA encoding a
CC tribonectin may be used in gene therapy. The present sequence represents
CC human MSF.
XX
XX
SQ Sequence 1404 AA:

Query Match 100.0%; Score 5698; DB 22; Length 1404;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1064; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKDNKKNTKKKKPPPPVNDAGSLDNGDKVTPPTSTTQHNKVTSEKITAKPIN 60
DB 200 VKDhkntrkkkppkpyvdaagsglndgkfvtlptsltglnkvstspkiltakpin 259
QY 61 PRPSLPKSDSKESLTVNKKETTKETTTNNKQTSDDKEKTSKKEQSIKTSK 120
DB 260 PRPSLPKSDSKESLTVNKKETTKETTTNNKQTSDDKEKTSKKEQSIKTSK 319
QY 121 DLAPTSLVAKPTPKAETTTGPAITTPKEPTPTPKPEASTTPKEPTTIKSAPTTPK 180
DB 320 dlaptskvlakptpkactttgpalittpkepttpkpeasttpkeptttiksaptttk 379
QY 181 EPAPTTTSAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 240
DB 380 epapttksapttkpeaptttkpeaptttkpeaptttkpeaptttkapttkpeaptt 439
QY 241 PKKAPPTPKPEAPPTPKPEPTTTPKEPAPTTKEPAPTTPKPEAPTTPKPEPA 300
DB 440 PKKAPPTPKPEAPPTPKPEPTTTPKEPAPTTKEPAPTTPKPEAPTTPKPEPA 499
QY 301 PTPKEPAPTTTKEPSPPTPKPEAPTTTSAPTTKEPAPTTTKEPAPTTTKE 360
DB 500 ptpkeaptttkespsptpkpeaptttkaptttkpeaptttkapttkkespspttk 559
QY 361 PAPTPPKPEAPPTPKKAPPTTTPKEPAPTTTTPKAPPAATPKKAPPAATPK 420
DB 560 paptpkpeaptttkkppaptttkpeaptttkkppaptttkkppaptttkk 619
QY 421 TPKKLTPTTPEKLAPTTPEKPAATPEELAPTTPEEPTTPEEPAATTPKAAAPNTPKE 480
DB 620 tpkkltpttpeklapttpekpaatpeelapttpeepttpeeepaptttkaaapntpk 679
QY 481 PAPTPPKPEAPPTPKPEAPTTTTPKGAATTPPKGAATTPPKKAPKELAPTTTK 540
DB 680 paptpkpeaptttkpeaptttketaptttkgtaptttkkppapkelaptttk 739
QY 541 EPTSTSDKPAATTPKGAATTPPKPEAPTTTTPKGAATTPPKGAATTPPKKAP 600
DB 740 eptstsdkpaattpkgaattpkpeaptttkkgaattpkgaattpkkaap 799

QY 601 KELAPTTTGGPTSTSDKPAATTPPKETAAPTTPKEPAPTTPKKAPPAATTPPEPTTSEVST 660
DB 800 kelaptttkgptstsdkpaattpketaapttpkpeaptttkkappaattppepttsevt 859
QY 661 PTTTKEPTTIHKSDESTPELSAEPPTKALENSKPEGVPTTKTPAATPKEMTTAKDKT 720
DB 860 ptttkcpttkhsdestpelsaepptkalenskppegvpttktpaatpkemttdakdt 919
QY 721 TERDLRTPEPTTAAPKMKETATTTETTESKITATATTOYSTTODTPEFKITTLTKT 780
DB 920 terdlrtpepttaapkmketatttetsteskitalatlvqsttdqtpfkittlkt 979
QY 781 TLAPKVTTKKTTITTTTEIMNKEEPAKPKDRATNSKATTPKOKPTAKPKKSTKKPKT 840
DB 980 tlapkvttkkttittteimnkeetapakdratnskattpokpptaakpkstkkpkt 1039
QY 841 MPVRKPKPTTPPKKMTSTMDLNTSLIAFAMLOTTTPROTNSKIVENPNSSEDAG 900
DB 1040 mpvrkpkpttpkkmstmdlntsliafamlottprotnskivenpnsxedag 1099
QY 901 AEGEPHMLLRPHVEMPEVTPDMOYLPRVNOGIIINPMLSDETNIGKRPVDGTTTLRN 960
DB 1100 aegephmllrphvempevtpdmoylprvnoqiilnplmsdetnigkryvdgttlrn 1159
QY 961 GTLVAFRGHYFWMLSPEFSPSPARITVEWGIIPSPIDTFTFRNCCEGKTFPFKDSQYWR 1020
DB 1160 gtlvafrghyfwmllspsfssparitvewgiipspidvtfrnccegkttffkdsqywr 1219
QY 1021 TNDIKDAGYPKPIKRGFGGLGQIYALSTAKYKMWPSVYEFK 1064
DB 1220 tndikdagypkplkrgfgglgqiyaalstakymwpsvyefk 1263

RESULT 5

AAB29778
ID AAB29778 standard; Protein: 902 AA.

XX AAB29778;

XX 28-FEB-2001 (first entry)

DE Human MSF-derived tribonectin.

KW Human tribonectin; MSF: megakaryocyte stimulating factor;

KW alternative splicing; joint boundary lubricant; O-linked oligosaccharide;

KW osteoarthritis; tribosupplementation; tissue adhesion inhibition;

KW friction coefficient reduction; gene therapy; antiarthritic;

OS Homo sapiens.

XX WO200064930-A2.

PD 02-NOV-2000.

PF 24-APR-2000; 2000WO-US10953.

PR 23-APR-1999; 99US-0298970.

PA (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.

PI Jay GD;

DR WPI: 2001-024673/03.

XX Novel tribonectin polypeptide useful as lubricant for treating

PT osteoarthritis, comprises O-linked lubricating moiety

XX Dislosure; Fig 1; 47pp; English.

CC The invention relates to a human tribonectin which is a product of
CC alternative splicing of the human MSF (megakaryocyte stimulating factor)
CC gene. The tribonectin has at least one O-linked oligosaccharide

XX Warman ML;
PI
XX
XX
DR WPI; 2001-182721/18.
XX
XX
PT New composition comprising the camptodactyly-arthropathy-coxa
PT vara-pericarditis protein in combination with an anesthetic, useful for
PT treating osteoarthritis, or as lubricants of tissue and joints -
XX
PS Example 1; Page -; 34pp; English

Query Match	100.0%;	Score 5698;	DB 22;	Length 1404;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1064;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

Oy	1	VHONKNNRKKKRPKPKPVYVDAAGSLDNGDKRVYTPDTSYTOHKNVSTSPKITTAKPIN	60
Db	200	Vdnnknmrkkkphcpkpvvdaagsldngddfvltprdstcqbhnkvsespklttakpin	259
Oy	61	PRPSLPNSDTSKNESTLVNKKETVYETKBTYTTNNKOTSDGKEKTTSAKETOSIEKTSK	120
Db	260	prpslpnsdtskneestlvnkketvctcttunnqtsldgkekttsaektsiektsk	319
Oy	121	DLAPTSKVLAKPAPKAETTTKGPALTTPKEEPPPTTPKEPASTTPKEEPPPTTKSAPTPK	180
Db	320	dlaptskvla kpkaetttk gpaltpkeppcttpkepastsctkpepcttkksapctpk	379
Oy	181	EPAPPTTTSKAPPTPKEPAPTTTKEBPATTPPKEPATTTKEBPATTTKSAPPTPKAPAPT	240
Db	380	epapcttkapctkpepacttkkepacttkkepacttkkepacttkksapcttkkepacttk	439
Oy	241	PKKPAPPTPKBPAPPTPKKEPPTTPPKKEPAPTTKKEBPATTPPKBPAPTPPKBPAP	300
Db	440	pkkpapctpkbpapctpkbpcttkkepacttkkepacttkkepacttkbpacttkbpakepa	499
Oy	301	PTTPKEBPATTTKEBSPPTPKBPAPTTTKSAPTTKEBPATTTKSAPTPKEBSPPTTKE	360
Db	500	pttpkebpacttkbpspttkbpapcttkksapcttkkepacttkksapcttkbpspttkke	559
Oy	361	PAPTPPKBPAPPTPKKBPAPTPPKBPAPTTTPPKBPATTTTKKAPAPAKBPAPPTPKETA	420
Db	560	papctpkbpapctpkbpcttkkepacttkkepacttkkbpacttkkepacttkpkeapcttk	619
Oy	421	TPPKUTPTTPPEKLAFTTPPEKPAPTTPEELATFTPEEPTTPPEBPAPTPPKAAAPNTPKE	480
Db	620	tpkktlftcpeklaftcpekpaftctpeelapcttpeaprtcttpeebpacttkaaapnctpk	679
Oy	481	PAPTPPKBPAPPTPKBPAPTPPKETAPPTPKGIAPTTLKEBPAPTPPKBPAPKELAPTTTK	540
Db	680	papctpkbpapctpkbpcttkkepacttkbpacttkkepacttkbpapcttkbpapkeapcttk	739

[illegible]

RESULT	4
AB29773	
ID	AB29773 standard; Protein; 1404 AA.
XX	
AC	AB29773;
XX	
DT	28-FEB-2001 (first entry)
XX	
DE	Human megakaryocyte stimulating factor (MSF), SEQ ID NO:1.
XX	
KW	Human MSF; megakaryocyte stimulating factor; triboneclin;
KW	alternative splicing; joint boundary lubricant; O-linked oligosaccharide
KW	osteoarthritis; tribosupplementation; tissue adhesion inhibition;
KW	friction coefficient reduction; gene therapy; antiarthritic;
XX	osteopathic.
XX	
OS	Homo sapiens.
XX	
PN	WO200064930-A2.
XX	
PD	02-NOV-2000.
XX	
PF	24-APR-2000; 2000WO-US10953.
XX	
PR	23-APR-1999; 99US-0298970.
XX	
PA	(RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.
PI	Jay GD;
XX	
DR	WPI; 2001-024673/03.
DR	N-PSDB; AAC81498.
XX	
PT	Novel triboneclin polypeptide useful as lubricant for treating
PT	osteoarthritis, comprises O-linked lubricating moiety -

PA (GENEY) GENETICS INST INC.
 XX
 PI Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K;
 XX
 DR WPI: 1992-284660/34.
 DR N-PSDB: MA027223.
 XX
 PT New human mega-karyocyte stimulating factors - for treating
 PT immune deficiencies, cancer, exposure to radiation or drugs,
 PT bacterial and viral infections, etc.
 XX
 PS Claim 1, 2 and 3: Fig 1: 87pp: English.
 XX
 CC The sequence given is a full length translation from the megakaryocyte
 CC stimulating factor (MSF) precursor. The sequence covered by exons II,
 CC III and IV encodes megakaryocyte stimulating factor (MSF). This
 CC sequence is modified by the addition of an N-terminal sequence encoding
 CC a secretory leader, an initiating methionine preceding exon II and a
 CC terminating codon following exon IV. The cDNA sequence given contains
 CC sequences derived from human megakaryocyte colony stimulating factor
 CC (meg-CSF). Exon I contains the initiating methionine, and encodes a
 CC classical mammalian protein secretion signal sequence. The sequence
 CC encoding the original meg-CSF includes exons II-IV and is thought to
 CC terminate in the region between amino acid residues 134 - 147. The
 CC primary transcript of this gene may be cleaved in different ways to
 CC yield a family of mRNAs each encoding a different MSF protein. Exons
 CC V and VI are thought to be related to the activity of the factor and
 CC are also implicated in the stability, folding and processing of the
 CC molecule. These exons are also thought to play a role in the observed
 CC synergy of MSF with other cytokines. Exons V and VI are believed to be
 CC implicated in the processing or folding of the appropriate structure of
 CC the resulting factor, i.e. one or more of these exons may contain
 CC sequences which direct proteolytic cleavage, adhesion, organisation of
 CC the cellular matrix or extracellular matrix processing. Both naturally
 CC occurring and non-naturally occurring MSFs may be characterised by
 CC various combinations of alternatively spliced exons from this sequence,
 CC with the exons spliced together in differing orders to form different
 CC members of the MSF family.
 CC
 XX
 SO Sequence 1404 AA:

Query Match 100.0%; Score 5698; DB 13; Length 1404;
 Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0;
 Matches 1064; Conservative 0; Indels 0; Gaps 0;

QY 1 VKDNKKRRTKKRPKPPVVDAGSGLDNGDFKVTTPDSTTQHNKYSTSPKITTAKPIN 60
 DB 200 vkdnknrtkkrrpprvvdeagsgldngdfvttptdsttqhknvstspkittakpin 259
 QY 61 PRPSLPNSDTSKESLTJVNKETVETKETTTNKTSTDGAKRTSAKTSOSIETSATK 120
 DB 260 prpslpnsdtsktsltvnketvettkttnkgtstdgakrtsaktsosietstak 319
 QY 121 DIAPTSVIAKPKPKAETTTKGPALTTPKPEPTTPPKKPASTTPKEPTTISAPPTPK 180
 DB 320 diaptsvialpkpkpkaetttkgsalttpkpepttpkkpastaipkpepttisapptpk 379
 QY 181 EPAATTTKSAPTTPKEPAPTTPKPAATTPKPAATTPKPAATTPKPAATTPKPA 240
 DB 380 epaapttksaptpkpeaptpkpaattpkpaattpkpaattpkpaattpkpaattpk 439
 QY 241 PKKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPA 300
 DB 440 pkkpaptpkpaattpkpaattpkpaattpkpaattpkpaattpkpaattpkpaattpk 499
 QY 301 PTPKKEPAPTTPKPEPTTPKPEAPTTPKSAPTTKSAPTTKSAPTTKSAPTTKSA 360
 DB 500 ptpkkepapttpkpepttpkpeaptpkpsapttksapttksapttksapttksapttke 559
 QY 361 PAPPTPKPAATTPKPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPA 420
 DB 560 papptpkpaattpkpkpaaattpkpaattpkpaattpkpaattpkpaattpkpaattpk 619

QY 421 TPKKLTPTTPEKLAAPTPEKPAATTPBEELAPTTPEEPPTTPEEPAPTTPKAAAPNTPK 480
 DB 620 tpkkltpttpeklaaptpekpaattpbeelapttpeeppttpeepaptpkkaapntpk 679
 QY 481 PAPPTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPA 540
 DB 680 papptpkpaattpkpaattpkpaattpkpaattpkpaattpkpaattpkpaattpk 739
 QY 541 EPTSTSDKAPAPTTPKGAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPA 600
 DB 740 eptstsdkapapttpkgaattpkpaattpkpaattpkpaattpkpaattpkpaattpk 799
 QY 601 KELAPTTKGPSTTSKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPA 660
 DB 800 kelapttkgspttskpaattpkpaattpkpaattpkpaattpkpaattpkpaattpk 859
 QY 661 PTTKKEPTTIHKSDESPELSAEPTRKALENSFKEGVPPTTKTPAATKPEMTTAKDKT 720
 DB 860 pttkkepittihsdespelsaeptrkalensfkegvpttktpaatkpemttakdk 919
 QY 721 TERDLRTPTTETTAAPKMTKSTATTETKTESKITATTQVSTTTODTTPPKITLKT 780
 DB 920 terdlrtpttettapkmstkstattettktteskitattqvstttodttppkitlkt 979
 QY 781 TLAPKVTTKTKTTTTEIMNKPEETAKPKDRATNSKATTPKPKPTKAPKPKPTSKKPK 840
 DB 980 tlapkvttktktttteimnkpeetakpkdratnskattpkpkptkapkpkptskkpk 1039
 QY 841 MPVRKPKPTTPPKRMISTMBELNFTSRIAPAMQTTRPNQOTNSKLYEVNPKSEBAG 900
 DB 1040 mpvrkpkpttpkrmistmbelnftsriapamqtrpnqotnsklyevnpksebag 1099
 QY 901 AEGTEPHMLLRPHVMPPEVTPDMOYLPRVPMOGIITNPMLEDETNICNGKRPDGLTTLRN 960
 DB 1100 aegtephmlrphvmppevtpdmoylprvpmogiiitnpmledeitnicngkpvdglttlrn 1159
 QY 961 GTLVAFRGHYEWMLSFSPSPARRITEVWGIIPSIDVFTFRNCCEGTFPKDSQIWR 1020
 DB 1160 gtlvafrghyewmlsfspsparritevwgiipsidvftfrnccegtffkdsqywr 1219
 QY 1021 TNDIKDAGYPRPIFKFGGLTGQIYAALSTAKYKRWMPSEVYFFK 1064
 DB 1220 tndikdagypripfkfggltgqiyaalstakyrwmpsevyffk 1263

RESULT 3
 AAB60568 standard; Protein; 1404 AA.
 ID AAB60568
 AC AAB60568;
 XX
 DT 27-APR-2001 (first entry)
 XX
 DE Human megakaryocyte stimulating factor (MSF, GACP).
 XX
 KW Human; GACP protein; camploactyl-arthropathy-coxa vara-pericarditis;
 KW MSF; megakaryocyte stimulating factor; synovial lubricant;
 KW chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;
 KW antiarthritic.
 KW
 OS Homo sapiens.
 OS
 PN WO200107068-A1.
 XX
 PD 01-FEB-2001.
 XX
 PE 21-JUL-2000; 2000MO-US20002.
 XX
 PR 23-JUL-1999; 99US-0145328.
 PR 19-JUL-2000; 2000US-0145328.
 XX
 PA (UYCA-) UNIV CASE WESTERN RESERVE.

DR N-PSDB: AAH98981.
 XX Isolated polypeptide for treatment of diseases, diagnostics, raising
 PT antibodies and research use -
 XX
 PS Claim 20: Page 1198-1201; 1275pp; English.
 XX
 CC The present invention provides the protein and coding sequences of novel
 CC proteins from a variety of organisms, including human, dog, cat, horse,
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
 CC from the organism of interest. They can be used in diagnostics,
 CC forensics, gene mapping, identification of mutations, to assess
 CC biodiversity and for nutritional purposes. The present sequence is a
 CC protein of the invention.
 XX
 SQ Sequence 1299 AA:

Query Match 100.0%; Score 5698; DB 22; Length 1299;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1064; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKDKKNNKTKKKPPKPPVVDGAGSLDNGDFKVTTPDSTSTQHNKYSTSPKITTAKPIN 60
 DB 200 vkdkknnrkttkkppkppvvdgagsgldngdfkvtcpdstctqhnkyslspkittakpin 259
 QY 61 PRESLPNSDTSKETSJLVNKKETTVEKETTNNKQSTDGKEKTSAKTOSIEKTSK 120
 DB 260 prslpnsdstsketsjlvnkkettvetkettlnkqstdgketsktsaketsiekttsak 319
 QY 121 DLAPTSTKVLAKPPPKAETTKGPAITPPKPTTPPKKPAATTPPKKPAATTPK 180
 DB 320 dlaptstkvlaakpppkatettkgpaitppkpttpkkpaattpkkpaattpk 379
 QY 181 EPAPTTTKSAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPT 240
 DB 380 epaptttksapttpkebapttpkebapttpkebapttpkebapttpkebapt 439
 QY 241 PKRPAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEB 300
 DB 440 pkrapapttpkebapttpkebapttpkebapttpkebapttpkebapttpke 499
 QY 301 PTPPKBAPTTPKEBAPTTPKEBAPTTPKSAPTTPKSAPTTPKSAPTTPK 360
 DB 500 ptpkbpapttpkebapttpkebapttpksapttpksapttpksapttpk 559
 QY 361 PAPTTPKEBAPTTPKPAATTPKEBAPTTPKPAATTPKPAATTPKPAATTP 420
 DB 560 papttpkebapttpkpaattpkebapttpkpaattpkpaattpkpaattp 619
 QY 421 TPKKLTPTPPEKLAAPTTPKPAATTPPEELAPTPPEEPPTTPPEAPTPPK 480
 DB 620 tpkkltptppeklaapttpkpaattppeelaptppeeppttppeaptppk 679
 QY 481 PAPPTTPKEBAPTTPKEBAPTTPKETAATTPKETAATTPKETAATTPK 540
 DB 680 pappttpkebapttpkebapttpketaattpketaattpketaattpk 739
 QY 541 EPTSTSDKAPATTPKGTAPTPKKEBAPTTPKGAATTPKGAATTPKGAATTP 600
 DB 740 eptstsdkapattpkgtattpkkebapttpkgaattpkgaattpkgaattp 799
 QY 601 KELAPTTKGPSTTSKAPATTPKETAATTPKETAATTPKPAATTPETPPT 660
 DB 800 kelapttkgpsttskapattpketaattpketaattpkpaattpetppt 859
 QY 661 PTTTKEPTTIHKSDESPBELSAEPTKALESKKEGVPTTKTAPATPKPMATTA 720
 DB 860 ptttkepttihsdespbelsaeptrkaleskkrgevppttktapatpkpma 919
 QY 721 TERDRTTPETTTAAPKMTKETATTEKTSKITATTOYTSNTTODTTEFFKITL 780
 DB 920 terdrttpettttaapkmtektattektskitatttoytsnttoddteffkitl 979

DB 920 terdrltpelettaapxkmtketatltekteskiltattqivstturgdtupfkitlkt 979
 QY 781 TLAPVTTKTKTTTITIMKKPEETAKPKRRAINSATTPPKPOKPPKAPKPTSTAKPK 840
 DB 980 tlapvttktktttittimkkpeetakpkrrainsattpkpkokppkapkptstakpk 1039
 QY 841 MPVRKRPKTPTPRKMSTWPELNPTRIAEAMLQTTTPRNPCKLVEVNPKRSFDAG 900
 DB 1040 mpvrkrpkrtptrkmtwpeelnptriaeamlqtttprnpcklvevnpskrsfdag 1099
 QY 901 AEETPMILRPHVFMPEVTPMDYLPVRNOCIIINPMLSDETNCNGKPYDGLTTLRN 960
 DB 1100 aeetpmlrphvfmpevtpmdylpvrnociiinpmldsctnngkpydglttlrn 1159
 QY 961 GTLVAERGHYFWMLSPPSPSPPARITEVWGLPSPIDVTFQNCNCGKTFPKDSQW 1020
 DB 1160 gtlvaerghyfwmlspspspparitevwglspidvtfqncncgktfpkdsqwr 1219
 QY 1021 TNDIKAGYPKPIFKFGGLTGQIVAAALSTAKYKKNPESVYEFK 1064
 DB 1220 tndikagypkpifkfggltgqivaalstakynknpesvyeffk 1263

RESULT 2

AAR26049
 ID AAR26049 standard; Protein: 1404 AA.

XX AAR26049;

DT 02-FEB-1993 (first entry)

XX MSF precursor.

XX Megkaryocyte colony stimulating factor; secretion signal; meg-CSF;
 XX stability; proteolytic cleavage; adhesion; alternative splicing.

OS Synthetic.

FT Key
 FT Location/Qualifiers
 FT 1..26
 FT /label= Exon_I
 FT 26..67
 FT /label= Exon_II
 FT 67..107
 FT /label= Exon_III
 FT 107..157
 FT /label= Exon_IV
 FT 157..200
 FT /label= Exon_V
 FT 200..1141
 FT /label= Exon_VI
 FT 1141..1166
 FT /label= Exon_VII
 FT 1166..1212
 FT /label= Exon_VIII
 FT 1212..1266
 FT /label= Exon_IX
 FT 1266..1331
 FT /label= Exon_X
 FT 1331..1373
 FT /label= Exon_XI
 FT 1373..1404
 FT /label= Exon_XII

XX W09213075-A.

XX 06-AUG-1992.

XX 17-JAN-1992; 92WO-US00433.

XX 18-JAN-1991; 91US-0643502.

XX 10-SEP-1991; 91US-0757022.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:16:16 ; Search time 107.17 Seconds
(without alignments)
735.411 Million cell updates/sec

Title: US-09-556-246-1_COPY_200_1263

Perfect score: 5698

Sequence: 1 VKDNKKRRTKKRTPKPYV.....VAALSTAKYKNPESVYFVK 1064

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq.1101.*
1: /SIDS8/gcgdata/geneseq/AA1980.DAT.*
2: /SIDS8/gcgdata/geneseq/AA1981.DAT.*
3: /SIDS8/gcgdata/geneseq/AA1982.DAT.*
4: /SIDS8/gcgdata/geneseq/AA1983.DAT.*
5: /SIDS8/gcgdata/geneseq/AA1984.DAT.*
6: /SIDS8/gcgdata/geneseq/AA1985.DAT.*
7: /SIDS8/gcgdata/geneseq/AA1986.DAT.*
8: /SIDS8/gcgdata/geneseq/AA1987.DAT.*
9: /SIDS8/gcgdata/geneseq/AA1988.DAT.*
10: /SIDS8/gcgdata/geneseq/AA1989.DAT.*
11: /SIDS8/gcgdata/geneseq/AA1990.DAT.*
12: /SIDS8/gcgdata/geneseq/AA1991.DAT.*
13: /SIDS8/gcgdata/geneseq/AA1992.DAT.*
14: /SIDS8/gcgdata/geneseq/AA1993.DAT.*
15: /SIDS8/gcgdata/geneseq/AA1994.DAT.*
16: /SIDS8/gcgdata/geneseq/AA1995.DAT.*
17: /SIDS8/gcgdata/geneseq/AA1996.DAT.*
18: /SIDS8/gcgdata/geneseq/AA1997.DAT.*
19: /SIDS8/gcgdata/geneseq/AA1998.DAT.*
20: /SIDS8/gcgdata/geneseq/AA1999.DAT.*
21: /SIDS8/gcgdata/geneseq/AA2000.DAT.*
22: /SIDS8/gcgdata/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5698	100.0	1299	22	AAAM24322 Human EST encoded
2	5698	100.0	1404	13	AAAR26049 MSF precursor. Sy
3	5698	100.0	1404	22	AAAB05058 Human megakaryoct
4	5698	100.0	1404	22	AAAB29773 Human megakaryoct
5	3484	61.1	902	22	AAAB29778 Human MSF-derived
6	1357.5	23.8	452	16	AAAB80041 Human megakaryoct
7	1168	20.5	5179	22	AAAM24516 C899P predicted am
8	1042.5	18.3	472	22	AAAB05059 Bovine MSF ortholo
9	981	17.2	763	21	AAAG38942 Arabidopsis thalia
10	950	16.7	1664	19	AAAW3106 C. thermocellum O1
11	771.5	13.5	1325	22	AAAM03645 Peptide #2327 enco

12	715	12.5	763	18	AAAM31852
13	625.5	11.0	4412	21	AAAY3666
14	612	10.7	572	18	AAAM31855
15	544	9.5	844	7	AAAB05070
16	542	9.5	807	21	AAAY54467
17	520.5	9.1	788	21	AAAY54466
18	508	8.9	1837	21	AAAB11726
19	506.5	8.9	744	9	AAAB2975
20	496.5	8.7	2971	21	AAAB41231
21	489.5	8.6	2972	22	AAAB50363
22	489.5	8.6	3118	22	AAAB50362
23	489.5	8.6	1721	21	AAAB11727
24	488.5	8.6	826	13	AAAR26042
25	488	8.6	617	22	AAAM16458
26	488	8.6	617	22	AAAM04187
27	488	8.6	957	21	AAAY9288
28	488	8.6	957	22	AAAM24513
29	488	8.6	1721	19	AAAR26042
30	485	8.5	1127	22	AAAB5541
31	476.5	8.4	652	9	AAAB2974
32	467.5	8.2	511	22	AAAM14883
33	467.5	8.2	511	22	AAAM27312
34	467.5	8.2	511	22	AAAM02607
35	450.5	7.9	378	12	AAAR14160
36	446.5	7.8	378	12	AAAR14162
37	446.5	7.8	750	20	AAAY05477
38	443	7.8	2870	21	AAAY5559
39	443	7.8	3178	21	AAAY5556
40	442.5	7.8	751	16	AAAR0839
41	433	7.6	2819	22	AAAB35408
42	428.5	7.5	2665	22	AAAM14533
43	428.5	7.5	2665	22	AAAM26950
44	428.5	7.5	2665	22	AAAM02259
45	428.5	7.5	3266	21	AAAB42491

ALIGNMENTS

RESULT 1	
ID	AAAM24322 standard: Protein: 1299 AA.
XX	AAAM24322;
AC	AAAM24322;
XX	
DT	12-OCT-2001 (first entry)
XX	
DE	Human EST encoded protein SEQ ID NO: 1847.
XX	
KW	Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW	tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW	diagnostics; forensic test; gene mapping; genetic disorder;
KW	biotechnology; gene therapy; nutrition.
XX	
OS	Homo sapiens.
XX	
PN	WO200154477-A2.
XX	
PD	02-AUG-2001.
XX	
PF	25-JAN-2001; 2001WO-US02687.
XX	
PR	25-JAN-2000; 2000US-0491404.
XX	
PR	17-JUL-2000; 2000US-0617746.
XX	
PR	03-AUG-2000; 2000US-0631451.
XX	
XX	15-SEP-2000; 2000US-0663870.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
XX	
PI	Cao Y, Drmanac RA, Zhang J, Wertman T,
XX	
DR	WPI; 2001-476164/51.

Mycobacterium tube
Sequence g1/101742
Mycobacterium tube
Sequence of the Fa
Amino acid sequenc
Amino acid sequenc
Cryptosporidium pa
Bioadhesive precu
Human ORFX ORF995
Human SRCAP. Homo
Human SRCAP. Homo
Portion of Cryptos
P. yoelii SSP2 ant
Peptide #2892 enco
Peptide #2869 enco
Human MUC11 polype
9900P predicted am
Cryptosporidium pa
Human protein sequ
Bioadhesive precu
Peptide #1317 enco
Peptide #1349 enco
Peptide #1289 enco
PRP 378. Triticum
PRP encoded by clo
C. albicans Rbt1 p
Caenorhabditis ele
Caenorhabditis ele
Japanese sea musse
Human 07CG27 gene
Peptide #967 enco
Peptide #987 enco
Peptide #981 enco
Human ORFX ORF2255

CC combinations of alternatively spliced exons from this sequence, with
 CC the exons spliced together in differing orders to form different
 CC members of the MSP family.
 XX
 SO Sequence 111 AA;

Query Match 8.3%; Score 613; DB 13; Length 111;
 Best Local Similarity 100.0%; Pred. No. 1.3e-26;
 Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 LSSCAGRCGEGYSRDATCNDYNCQHWECCPDFKRVCTAELSCGRCFESFERGRECD 62
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 8 LSSCAGRCGEGYSRDATCNDYNCQHWECCPDFKRVCTAELSCGRCFESFERGRECD 67
 QY 63 DAOCRRYDKCCPDYSPCAEVHNPSPSSKKAPPSSGASOTIK 106
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 68 daqckkydkcepdysfcaevhnpstppsskkappssgasqtlk 111

Search completed: April 26, 2002, 16:16:03
 Job time: 154 sec

E-1.4

```

OY 43 ELSCKGRCFES-----FERGECDDACCKKYDKCCPDYESFCA-----EYHNTPSPSS 92
DB 3278 qlschvggsepilrlqwlkagire-----lkpsdrctsfisagtalrlrdvakadsgdyv 3331
OY 93 KKAPPPGASQOTIKSTTKRSPKPPNKKKKKYESEE---ITEEHVSSENOESSSSSSSS 149
DB 3332 ckaasnvag-sdttskvtikdkpavapatkkaavagrlffvsepgslvvektatlfiak 3390
OY 150 SSSSTIMIKSSKNSANREL-----QKLIKVDKNKNRTKKKPKPPYVD 196
DB 3391 vggdplpnvkwtkgkw---rlqngagrvflhqgdeaklfdltklt----- 3434
OY 197 EAGSLDNGDKRVTPDSTTOHKNVSTSPKTTAK-----PIMRPSLPN 243
DB 3435 -----dsqlyrc-----valfhegheisenvulqvderrkkqekiegdlamllktrpilkkg 3484
OY 244 SDTSKETSILTVKKEETTVKEKETTNNKOTS--TDGKEKTTSAKETSISIEKT-----SAK 295
DB 3485 ageeeelal-mellknvdpkveyekyarmygidfgrllqrvgeehrvckhvievfeae 3543
OY 296 DLAPTSKVLAKPJPRAETTTKGPALT-----TPKEPTPTP--KEPASTTPKEPTPTTKS 349
DB 3544 evevfevkapp-----kgpelsekllppktpkvpkpkppakvpevkkivvee 3595
OY 350 APTTKREP-APTITTSAPTTKEPAPTITTKEPAPTITTKEPAPTITTKSAPTTP 408
DB 3596 kvrvpeevpplkvpevlppkevvp-----ekkvvpakppea-----ppkvp 3641
OY 409 KEPAPTTPKK-----PAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP-----KE 457
DB 3642 eapkevpevpekkvppppkpevpplkvpevkaavpekkvpealppkpsppvefeep 3701
OY 458 PAPAPAKKAPPTTPKEPAPTITTKEPAPTITTKESPPTTPKEPAPTITTKSAPTTPKEPAPT 517
DB 3702 espsappp-kpevpvrvpevpevpevpevpevpevpevpevpevpevpevpevpevpev 3757
OY 518 TKSAPTTPKEPSPPTTKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 577
DB 3758 kvvpvp-ppkpkpevpplk-----vpevkvavpevpevpevpevpevpevpevpevpev 3807
OY 578 PTAPKEPAPTTPKEPAPTTPPKKLT-----PTTPKLTAPT-----PEKPAPTTPEBELAPTTP 629
DB 3808 evalpeepveevpeepaapqvtvppknpvpekkapavavakpelpkvpevpevpevpev 3867
OY 630 EEPPTP--TTPKEPAPTTPKKAAPNTKEPAP-----TTPKEP-----APTTPKEPAPTTPK 678
DB 3868 ekvplvvpkk--peapppakvpevpevpevpevpevpevpevpevpevpevpevpevpev 1 3921
OY 679 ETAPTTPKGTAPTTLKEPAPTTPKPKAPKELAPTITTKESTSTSDKAPAPTTPKGTAPTTP 738
DB 3922 eekpavp---vperaesppvevye---peeiap-----eeelapeekpvpvae---eep 3969
OY 739 KEPAPTTPKEPAPTTPKGAAPTTLKEPAPTTPKPKAPKELAPTITTKGPTSTSTSDKAPAPT 798
DB 3970 evpppavpeepkllpekkvvp-vlkkpkaapppkepekev---lekpllkprpppppap 4025
OY 799 PKETA-----PTTPKEPAPTTPKKP--APTTPETPTPTSEVS--TPTTKEP----- 842
DB 4026 pkevdevkeklfqlkaipkpkvpenpvyekveltlpklvpggekkvrvllpeirkepeveev 4085
OY 843 --TTIKHSDESTPELSAETTPKALENSPKPEGVPTTKTPAANKP-----EMTTAKDKTT 896
DB 4086 lksvllkkrpeeeepkve---pkkle-kvkkpavp---eppppkpvveevpvlvltkrex 4137
OY 897 ERDLKRTPTETTTAAPKMTKATATTTTEKTTSKITATTTTQVTSSTTQDTPTFFKTTTKTT 956
DB 4138 lpeptkvpeklpailpapepdkpkea-----evklkppp 4173
OY 957 LAKVTTTKKTTITTTTEIMNK-----PEETAKPKDRATNSKATTPKQKTPKPKPTS- 1009
DB 4174 veepepfpiaavlvpyvgkkaeakapkeekakpkipikvypkktptpiseerklklp9sg 4233
OY 1010 -TKRPMTPEVRKPKTTPP-----RKMTSTMBELNETSRIAEAMLOTTP 1053

```

```

DB 4234 gekpdpdeapfryqlkavplkfvkeikdlilesefvgssalfecfvspst 4283
RESULT 15
AAR26050
ID AAR26050 standard; Protein: 111 AA.
AC AAR26050;
XX
XX
XX 02-FEB-1993 (first entry)
XX
XX MSF-K130.
DE
DE Megakaryocyte colony stimulating factor; secretion signal; meg-CSF;
KW stability; proteolytic cleavage; adhesion; alternative splicing;
KW MSF-precursor.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
FH Region 1..4
FT Region /label= Enterokinase_site
FT Region 7..48
FT Region /label= Exon_II
FT Region 48..88
FT Region /label= Exon_III
FT Region 88..111
FT Region /label= Exon_IV
FT
FT WO9213075-A.
PN
PN 06-AUG-1992.
PD
PD 17-JAN-1992; 92WO-US00433.
XX
XX 18-JAN-1991; 91US-0643502.
XX 10-SEP-1991; 91US-0757022.
XX
XX (GEMV ) GENETICS INST INC.
XX
XX Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K;
PI WPI; 1992-284660/34.
XX
XX N-PSDB; AAQ27224.
DR
DR
XX
XX New human mega-karyocyte stimulating factors - for treating
PT immune deficiencies, cancer, exposure to radiation or drugs,
PT bacterial and viral infections, etc.
XX
XX Disclosure; Fig 3; 87pp; English.
XX
XX The sequence covered by exons II, III and IV encodes megakaryocyte
XX stimulating factor (MSF). This sequence was used to produce MSF as
XX a fusion protein with thioredoxin. This sequence was derived from a
XX MSF-precursor (see also AAR26050).
XX
XX The precursor cDNA sequence contains sequences derived from human
XX megakaryocyte colony stimulating factor (meg-CSF). Exon I contains
XX the initiating methionine, and encodes a classical mammalian protein
XX secretion signal sequence. The sequence encoding the original meg-CSF
XX includes exons II-IV and is thought to terminate in the region between
XX amino acid residues 134 - 147. The primary transcript of this gene
XX may be cleaved in different ways to yield a family of mRNA's each
XX encoding a different MSF protein. Exons V and VI are thought to be
XX related to the activity of the factor and are also implicated in the
XX stability, folding and processing of the molecule. These exons are
XX also thought to play a role in the observed synergy of MSF with other
XX cytokines. Exons V - XII are believed to be implicated in the
XX processing or folding of the appropriate structure of the resulting
XX factor, ie. one or more of these exons may contain sequences which
XX direct proteolytic cleavage, adhesion, organisation of the cellular
XX matrix or extracellular matrix processing. Both naturally occurring
XX and non-naturally occurring MSF's may be characterised by various
XX

```

XX Claim 5; Fig 13; 55pp; English.
 PS
 CC This novel 74 kDa protein is encoded by an open reading frame of
 CC a Mycobacterium tuberculosis DNA fragment (see AAY3610) containing
 CC polymorphic GC-rich sequences. Its amino acid sequence shows
 CC a high proline content, but there is no homology to any known
 CC proline-rich antigens of mycobacteria. Novel M. tuberculosis
 CC proteins (see AAY1851-57) are claimed. These can be produced as
 CC recombinant proteins, especially in bacterial, yeast, fungal or
 CC higher eukaryote host cells, and used for diagnosing tuberculosis
 CC and other mycobacterial infections in humans or animals. The
 CC claimed proteins can also be used for epidemiological studies, for
 CC monitoring vaccination, and for the development of vaccines and
 CC anti-mycobacterial drugs.
 CC
 XX
 SO Sequence 763 AA:

Query Match 9.6%; Score 715; DB 18; Length 763;
 Best Local Similarity 28.2%; Pred. No. 3.1e-31;

Matches 248; Conservative 39; Mismatches 361; Indels 232; Gaps 39;

```

OY 307 PPKKATTTTGGALTPPKKPTPTTPKKEPASTTPKKEPTPTTTSAPTTPKKEPAPTTPKSAP 366
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 3 PVP-----apralaplpappapapapkskpfppap-----pappcmwlvsaap 46
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 367 TTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKSAPTTPKKEPAPTTPKKEPAPTTPKKE 426
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 47 pcp--pap-----pappkkskafppvpapparelapplp--pap-----paape 90
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 427 PAPTTPKKEPTPTT---PKKAPTTPKKEPAPTTPK--EPAPTAPK--KPAPTTPKKEPAPTTPK 480
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 91 srpalppcpvpvlpdppeapapvpapnsppfpfpfpapkfvpapvp--pvpnspp 148
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 481 EP--APTTPKKEPAPTTPKKEPAPTTPKKSAPTTPKKEPAPTTPKKSAPTTPKKEPAPTTPK 538
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 149 fppfppeaalnpap-----pappianspplppappipaglt--ppaapwvpaapaskpa 201
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 539 TTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKE 594
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 202 sprrppap-----pmpapfmeffpfpvpdpplsketppapppappppvpplppvpplp 256
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 595 TTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKE 644
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 257 pvpnkllpppp---appvaavaavlpapcpplpplpnnpappapppvpplpplpnshp 313
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 645 ---PKAAAPNTPKKEPAP----- 658
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 314 pappapavpvgvlpaplpisgrvsvwksftlslfcccvcsgewlagalnpssrspl 373
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 659 --TTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKE 708
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 374 tttcpalpapplpplpplpintavpplpplpvtalappplpplpplpplpispvgvpap-- 431
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 709 LAPTTTKEPTSTSDKAPPTTPKKGTAAPTTPKKEPAP-----PTTPKKEPAPTTPKKGTAAPTTL 762
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 432 --plppgkfwltpplappapepk--lvpylpvgpscpssekpppapppepbebskpsalpp 488
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 763 KEPAFTTP---KKPAPKELAPTTTGGP--TSTSDKAPAPTTPKET---APTTPKKEPAPTTP 815
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 489 appapssmsavrvpssplpappapaprasmpalppappspatllcpplppspapnsp 548
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 816 KKPAFTTETETPTTSEVSTPTTKEPTTIHKSP---DESTPELSAEPTPKA--LENSP 869
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 549 --pappapptcpklls--anppcpvpvapnppappapppapppelppapdpplpvpansp 604
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 870 KEGVPTTKTPATKPEMTTTPAKDKTERDLRTPTETTTAPKMKKEATTEKTESKI 929
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 605 --pappappppsalpfvnpa-----pplpaapk-----srpal 637
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 930 TATTQVSTTTQDTTPPKITTLKTTTLAPKVTTTKKTTTTEIMNKPEETAKPKDRATN 989
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

DB 638 paappappppvratrp-----ppappappapn 665
OY 990 SKATTPPKPOKP---TKAPKKPTSTKKPKTMPVRRKPTTPPKMTSTM-----PELNP 1040
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 666 smalpappdpplppllatppappappplpmppappplpppaappppplpnpappplp 725
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 1041 TSRIAEMNQTTTRPMPQNTNSKLYENPNPSEDDGAGEGCT 1080
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 726 vpgaplaplpingrpvfarfarknsll-----gsssgdt 756
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 14
 AAY53666
 ID AAY53666 standard; Protein; 4412 AA.
 XX
 AC AAY53666;
 XX
 DT 22-FEB-2000 (first entry)
 XX
 DE Sequence gi/1017427/emb/CAA62189 from an alignment with protein 608.
 XX
 KW Mechanical stress; gene therapy; protein 608; osteoporosis; bone density;
 KW bone development; gi/1017427/emb/CAA62189.
 XX
 OS Unidentified.
 XX
 PN WO9960164-A1.
 XX
 PD 25-NOV-1999.
 XX
 PF 14-MAY-1999; 99WO-US11066.
 XX
 PR 15-MAY-1998; 98US-0085673.
 XX
 PA (QDAR-) QDARK BIOTECH INC.
 XX
 PI Elnat P, Mor O, Skallter R, Feinstein E, Faerman A;
 XX
 DR WPI: 2000-053304/04.
 XX
 PT Identification of stress induced genes for determining risk and
 XX
 PS preventing, treating or controlling osteoporosis
 XX
 PS Claim 32; Fig 6A-R; 308pp; English.

The present sequence is obtained from a clustral X alignment with protein 608. Protein 608 was identified using the method of the invention after subjecting rat osteoblasts to mechanical stress. Expression of the 608 gene was found to be upregulated by about 3-fold in cells subjected to mechanical strain. The specification describes a method for the identification of genes responsive to a specific mechanical stress. The method comprises applying the mechanical stress to an organism (tissue or cells comprising bone cells), isolating the specific cellular fractions and extracting mRNA from them, and differentially analysing the mRNA in comparison with control samples. The method is used to identify genes whose expression is responsive to a specific stress. The identified genes are employed in determining risk associated with a physiological or disease state. The risk determination methods are used for testing a medication for gene therapy. These medications, or genes identified by the method of the invention, are used for treating, preventing or controlling a physiological or disease state (especially osteoporosis or bone density) or other factors causing or contributing to osteoporosis or its symptoms or other conditions involving in mechanical stress or its lack. The methods can also be used for advancing research or studies in bone development.

Sequence 4412 AA:

Query Match 8.7%; Score 645.5; DB 21; Length 4412;
 Best Local Similarity 24.0%; Pred. No. 1.3e-26;
 Matches 271; Conservative 128; Mismatches 488; Indels 243; Gaps 52;

PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140891.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142820.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.

PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148584.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149829.
PR 23-AUG-1999; 99US-0149802.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157573.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160747.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.

QY		836	PRTTKEPTTHKSPDESTPELSAPPTPKALENSPKREGVATTTTPAATKPEMTTAKDKT	895
Dd		590	-pttkepattkkepapt---kepapt---tkkep-apttkkepapttkkepaptke---	636
QY		896	TERLRRTPPETTTAAARKWTKETATTTKTETESKITATTTOVTSTTTODTTFPKITTLKTT	955
Dd		637	-----papt-----tkkepapt-----kepapttkkepapttkkepap-----	667
QY		956	TLABKVTTTTKTTTTTEIMNKPEETAAPKDRAINSKATTTTRPOKPFAKPKPTSTKKPPT	1015
Dd		668	-----tkkepapttkkep-apt-tkep--	686
QY		1016	MPRVRKPTTTPPPRKMSTMPBELNPTSRIAEAMLQTTPRNQPNPSKLTVNVNPKSEDAG	1075
Dd		687	-----apltclptkrmtstmpelnpustliaeaml-tttprnqpnslalvevnpksedag	739
QY		1076	AEGEPHALLRPHYFMPEVTPDMDYLPRVPNOGIIINPMLSDETNIENGKRPVDGLTTLRN	1135
Dd		740	aeegepnmlllpvhyfempetvpdmaylprvpngjlllnpmisdetenlcngkpvqdltelrn	799
QY		1136	GTLVAFRGHYTEMMLSPESPSPARRITVWGIIPSIDVTFTRCNECKTFFPKDSOYWR	1195
Dd		800	glivafirghyfmmlspisfparsparillewvgipspidvtftcrceegkltfikdsywrf	859
QY		1196	TNDIDAGYPPRIEFGFGLNGOIYAALSTAKYKNWPESVYEFK	1239
Dd		860	tndldagypkrlfkgrlg9lbgqlvaalsta-yknwpesvyffk	902
RESULT 6				
AAAR80041	ID	AAAR80041 standard; Protein: 452 AA.		
XX	AC	AAAR80041;		
XX	DT	10-APR-1996 (first entry)		
XX	DE	Human megakaryocytopoientin protein.		
XX	KX	Human; megakaryocytopoientin; wheat germ agglutinin; heparin;		
KW	KW	megakaryocyte; aplastic anaemia; thrombocytopenia; hematocytopenia;		
KW	KX	multipotential stem cell.		
OS		Homo sapiens.		
XX	FH	Key Location/Qualifiers		
FT	FT	Misc-difference 393..396 /note= "unspecified amino acids"		
FT	FT	Misc-difference 444..446 /note= "unspecified amino acids"		
XX	XX	WO9523861-A1.		
PN	XX	08-SEP-1995.		
PD	XX	06-MAR-1995; 95MO-CN00015.		
PF	XX	04-MAR-1994; 94CN-0112066.		
XX	XX	(SHAN-) SHANGHAI BEITI BIOTECHNOLOGY CO LTD.		
PI	XX	Gu X, Han Z, Shen Q:		
XX	XX	WPI; 1995-320576/41.		
DR	XX	N-PDB; AAT04546.		
PT	XX	New haematopoietic cell growth factor - used for treating		
PT	XX	thrombocytopenia and hematocytopenia		
PS	XX	Example; Page 23; 36pp; Chinese.		
CC		This sequence represents the human megakaryocytopoientin (MPO) protein.		

This sequence was purified using a carrier which can couple wheat germ agglutinin and heparin to separate MPO⁻. Fragments of this sequence (see AAB80039 and AABR0040) were used to produce the amplification primers shown in AAT04544 and AAT04545. The fragments amplified by these primers can then be used as probes to screen human cDNA libraries for MPO cDNA.

The MPO cDNA can then be inserted into a plasmid which is used to transform cells to produce MPO. The MPO sequence is capable of promoting colony formation of megakaryocytes, enlarging the size of megakaryocytes and stimulating the proliferation of multipotential stem cells. The factor may be used for treating thrombocytopenia and hematocytopenia.

The purification method can be used to isolate MPO from human urine or sera of patients with aplastic anaemia, and from animal blood or urine by radiation exposing the animals to induce aplastic anaemia.

Query Match	28.1%	Score 2081	DB 16	Length 452
Best Local Similarity	83.5%	Pred. No. 9.3e-106		
Matches 395	Conservative 3	Mismatches 21	Indels 54	Gaps 2
QY	1	ODLSSCAGRCGEYSYSDAHCNCDYNCQAHMECCPDEKRYCAETALSCGRCGFESEFGRGC	60	
DB	25	qdlsscagrcgcygysrdalcnocynqhmecpdlkryctaelascgycfsefergrec	84	
QY	61	DCDAGCKRTDKCCPDYEFRCAEVHNPTSPSSKAPPPSGASOTIKSTKRSPKPNKKK	120	
DB	85	dcdagckkydkccpdyefrcaevhmptspsskappsgasotlksrskppnkkk	144	
QY	121	TKKVISEETTEEHSVSENOESSSSSSSSSSSTIKIKSKNSAANRELQKKLVKDK	180	
DB	145	tkkviaseelte-----vkdnk	161	
QY	181	KNRTKKKPTPKPPVVDAGSLDNGDFKVTPTSTQHNKSTSPKITAKPILNRPPL	240	
DB	162	knrtkkkppkpvpvvdagsglndgfkvtptstqhnkvspskitakpilmprsl	221	
QY	241	PPMSDTSKETSLVYNKETTVETKETTNTQSTNDGKEKTTSAKETQSILEKTSADLAPT	300	
DB	222	ppmsdtkselslvnketvtetktettntqstndgkelttsaketqsaktdlapt	281	
QY	301	SKVLAKPTPKAETTTKGPAITLTKPEPTPTPKPEASTTKPEPTPTTKSAPTPKPEAPT	360	
DB	282	skvlakptpkpaetltkgpaitltpkeptptpkpeasttkpeptptlksaptpkpeapt	341	
QY	361	TTTSAPTPKPEAPTPTTKPEAPPTPKPEAPPTTKPEAPTTTKSAPTPKPEAPTTPKPA	420	
DB	342	ttksaptpkpeappttkpeapptpkpeappttkpeappttkshprlscxxxxcqp-	400	
QY	421	PTTPKEBAPTPPKPEPTPTTKPEAPPTTKPEAPPTPKPEAPTPAKKAPPTPKPE	473	
DB	401	pttpkebaptppkpeptpttkpeappttkpeappttkpeappttkshprlpe	443	
RESULT	7			
AAB60569				
ID	AAB60569	standard; Protein: 472 AA.		
XX				
AC	AAB60569;			
XX				
DT	27-APR-2001	(first entry)		
XX				
DE	Bovine MSR orthologue, superficial zone protein (SZP).			
XX				
KW	Bovine; CACP protein; campodactyly-arthropathy-coxa vara-pericarditis;			
XX				
KW	superficial zone protein; SZP; MSR orthologue; synovial lubricant;			
XX				
OS	osteoarthritis; joint lubrication; osteopathia; anilarthritis.			
XX				
OS	Bos taurus.			
XX				
PN	W0200107068-A1.			
XX				
PD	01-FEB-2001.			


```

|||||
Db 865 eptlhkspdestpelsaepfklenspkpvgpvtlcpaatkpcemtltaakdktererl 924
QY 901 RTTPPTTTAARKMTKETATTTTEKSTTATTOVTSTTODTTPETTKTTTLAK 960
Db 925 rttepetlaapkmktetatektektskltatctqstlctdtpckltlcltlak 984
QY 961 VTTTKTTTTEIMNKPEETAKPKDRATNSKATTPKQKPTAKKPTSTKKPTMPVR 1020
Db 985 vtctkctlttelmkpeetaktakdratskaltcpkqpktpakkkpstatckpktmpvr 1044
QY 1021 KKTPTTPPKMTSTMBELNPTSRIAPAMQTTRPNQTNKSLIVENPKSEDAAGACGET 1080
Db 1045 kptctptckmtstmbelnpstriaeamqcttrpnqtpnkskivevnpksedaggaaget 1104
QY 1081 PHMLARPHFMEVPEVDMVLPVPVNOGIIINPMLSDETINCGKRVDDLTLRNGTLVA 1140
Db 1105 phmlarphfmevpcpmvlpvpvnggiiinpmisdenicngkpvdgdlctlingtlva 1164
QY 1141 FRGHYFMMLSPPSPSPARITEWGISPIDVTFRNCCEGKTFEPKOSQKWRFTNDIK 1200
Db 1165 frghyfwmlspfsparritewgispidvtfrnceogkttffkdsqywrftndik 1224
QY 1201 DAGYPRPIKRGEGGLTGOVVAALSTAKYKNMPEVYEFKRGSGTGOYIKOEPRVCKPGR 1260
Db 1225 dagypkpiKgi9glcqlvaalstakympevyfkrigsigqiykqdepvqkcpgr 1284
QY 1261 RPLANTPVGEMTOVRRRFRERAIGPSQHTTIRIQSPARLAYQDKGLHNEVKSILMR 1320
Db 1285 rplantpvgemtqvrrrrteraligsqhtlirigsparlayqdkgvlhnevksilmr 1344
QY 1321 GLPNVTSALSLPNIRKPDGIDYAAFSKDOYVINDVPSRARAITRRSGOTLSKVMYNCP 1380
Db 1345 glpnvtsalslplnrkpdgidyaafskdgyndvpsraraaitrrsgotlskvmyncp 1404

RESULT 4
ID AAM24322 standard; Protein; 1299 AA.
XX AAM24322:
XX
XX AAM24322:
XX
XX 12-OCT-2001 (first entry)
XX
XX Human EST encoded protein SEQ ID NO: 1847.
XX
XX Human: sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
XX tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
XX diagnostics; forensic test; gene mapping; genetic disorder;
XX biodiversity; gene therapy; nutrition.
XX
XX Homo sapiens.
XX
XX WO200154477-A2.
XX
XX PD 02-AUG-2001.
XX
XX PF 25-JAN-2001; 2001WO-US02687.
XX
XX PR 25-JAN-2000; 2000US-0491404.
XX PR 17-JUL-2000; 2000US-0617746.
XX PR 03-AUG-2000; 2000US-0631451.
XX PR 15-SEP-2000; 2000US-0663870.
XX
XX (HXSE-) HXSE0 INC.
XX
XX Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
XX Cao Y, Drmanac RA, Zhang J, Werhman T;
XX WPI; 2001-476164/51.
XX DR N-PSDB; AAH98981.
XX
XX Isolated polypeptide for treatment of diseases, diagnostics, raising

```

```

PT antibodies and research use
XX
XX Claim 20; Page 1198-1201; 1275pp; English.
PS
XX
XX The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a
CC protein of the invention.
XX
SQ Sequence 1299 AA;
Query Match 92.2%; Score 6834; DB 22; Length 1299;
Best local similarity 100.0%; Pred. No. 0;
Matches 1271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 QDLSSCAGCGGSGYSRDANCNDYNOHMECCPPKRYCTAELSGRCFSEFERGEC 60
Db 25 qdlsscaagrcggystrdancndynohmccppdkryctaelsgrcfsefergrec 84
QY 61 DCDAGCKKYDKCCPDYSEFCAEYHNPTSPSSSKAPPPSGASQTIKSTTKSPKPNKK 120
Db 85 dcdagckkydkccpdysefcaeahnptspssskappsgasqtlkstkspkpnkk 144
QY 121 TKKVTESEITTEHSHVSENOESSSSSSSTTIWKIKSSKNSANRELQKKIWKDK 180
Db 145 tkkvteeseittehshvseenoesssssssttiwkikssksanrelqkklwkdk 204
QY 181 KNRTEKKKPPKPVVDAGSGLDNGDFKVTTPDTSTYQHNKYSTSPKITANKPINRPSL 240
Db 205 knrtekkkppkpvvdaagsgldngdfkvttpdststynknkystspkitankpinrpsl 264
QY 241 PPNSDTSKETSILTVNKEETVEKETTNNKQSTDSKEKTTSAKETQSIEKSANLAPT 300
Db 265 ppsndtsketsilvnrketvetttnnkqstdskektttsaketqsieksanlapt 324
QY 301 SKVLAKPTKAEFTTGGPALTTKPEPTPTPKRPASTTPKREPTTIKSAPTPKRPAPT 360
Db 325 skvlakptkaefttggpalttkpeptptpkrpasttpkrepttiksaptpkrpapt 384
QY 361 TTKSAPPTKPEAPPTTKPEAPPTTKPEAPPTTKPEAPPTTKSAPPTKPEAPPTPKKA 420
Db 385 ttksapptkpeappttkpeappttkpeappttkpeappttksapptkpeapptpkka 444
QY 421 PTPPKPEAPPTTKPEAPPTTKPEAPPTTKPEAPPTTKPEAPPTTKAKKAPPTPKRPA 480
Db 445 ptppkpeappttkpeappttkpeappttkpeappttkpeappttkakkapptpkrapa 504
QY 481 EPAPTTKEPSPTPKPEAPPTTKSAPPTTKPEAPPTTKSAPPTTKPEPSPTTKPEAPPT 540
Db 505 epapttkpesptpkpeappttksappttkpeappttkpeappttkpepspttkpeappt 564
QY 541 PKPEAPPTPKRPAPTTKPEAPPTTKPEAPPTTKKAPAPAPKPEAPPTPKETAPPTPKKL 600
Db 565 pkpeapptpkrappttkpeappttkpeappttkkapapapkppeapptpketapptpkkl 624
QY 601 TPTPKELAPPTPKKAPAPPTPEELAPPTPEEPPTTPEEPAPPTPKKAAANTPKPEAPPT 660
Db 625 tptpkelapptpkkapapptpeelapptpeeppttpeepapptpkkaaantpkpeappt 684
QY 661 PKPEAPPTPKPEAPPTPKETAPPTPKGTAPTTLKPEAPPTPKKAPKELAPPTTKEPTST 720
Db 685 pkpeapptpkpeapptpketapptpkgtapttlkpeapptpkkapkelappttkepst 744
QY 721 TSDKPAPTPKGTAPTTKPEAPPTTKPEAPPTTKGTAPTTLKAPATTTKKKPKELAP 780
Db 745 tsdkpaptpkgtapttkpeappttkpeappttkgtapttlkapeappttkkpkapelap 804
QY 781 TTTGCPSTSTSDKRAPPTPKETAPPTTKPEAPPTTKKAPAPAPPTTPEPTTSEVSTPTTK 840

```

QY	1081	PHMLLRPHVPEVTPDMQDYLPRVPMQGIINPMLSDETNTICGAPVDGTLTTLNGELVA	1140
Db	1105	phmlrlprhvipevtpdmqdlprvpmqgillnplmsdetnticngxpvqgltcllntngelva	1164
QY	1141	FRGHYFWMLSFSPSPARRTITVMGIPSPIDPVPFRCCCEKTEPFEDSOYMFRTMDIK	1200
Db	1165	fgrhyfwmlsfpspsparrititvmgipspidpvpfrcccektelfidsaywrtmdik	1224
QY	1201	DAGYPRKIRIFGFGGLTGQYVALSTAKKYNPESYVFEKRGSSIQOYTYKGPVOKCPGR	1260
Db	1225	dagypkririfgfggltgqyvalstakkympesyvfekrsgsiqylykqevpqkcpgr	1284
QY	1261	RPALNYPVYEGEMQVRRRRRERAIIGCSQHTIRIYSPARLAYOKGVLHNEVVSILMR	1320
Db	1285	rpalnypvyegemqvrerrrreraiigscqhtlrlqysparlayqdkyjlhnevvsilmr	1344
QY	1321	GLPVMVTSALSLNIRKPPGCOYVAFSPKSOQYINIDVPSRTKRAITTRSGOTLSKVMYNCP	1380
Db	1345	glpvmvtsalslnirkippgcoyvafspksqyindvpsrtkraittrsgotlskvmyncp	1404

RESULT	3
AAB29773	
ID	AAB29773 standard; Protein: 1404 AA.

DT 28-FEB-2001 (first entry)

Human megakaryocyte stimulating factor (MSF), SEQ ID NO:1.

KW Human MSF: megakaryocyte stimulating factor; fibrinectin;
KW alternative splicing; joint boundary lubricant; O-linked oligosaccharide;
KW osteoarthritis; fibrinolytic; tissue adhesion inhibition;
KW osteoclast coefficient reduction; gene therapy; antiarthritic;
KW osteopathic.

Homo sapiens.

PN WO2000064930-A2.

02-NOV-2000. PD

PF 24-APR-2000; 2000WO-US10953.

PR 23-APR-1999; 99US-0298970.

PA (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER
VY

PI Jay GD;
xy

DR WPI; 2001-024673/03.
DR N-PEND; 2000-01408

XX	DT	Novro] + + + 1 b o o o + 1 +
----	----	------------------------------

PT osteoarthritis, comprises O-linked lubricating moiety -
XY

PS Claim 3; Page 7; 47pp; English.
xy

The invention relates to a human tribonectin which is a product of alternative splicing of the human MSF (mesenchymal cyto stimulating factor) gene. The tribonectin has at least one O-linked oligosaccharide lubricating moiety and has a polypeptide sequence comprising 1-76 repeats of a motif having at least 50% identity to the sequence KEPAPT (AA829774). The invention also relates to a nucleic acid encoding a human MSF-derived tribonectin; a biocompatible composition comprising a human tribonectin for inhibiting tissue adhesion formation; and a method of diagnosing osteoarthritis or a predisposition to osteoarthritis by measuring the amount of MSF or its fragment in a biological sample of a mammal, wherein an increased amount of MSF compared to a control indicates the presence of or predisposition to developing osteoarthritis. The tribonectin and DNA encoding it are useful in the treatment of osteoarthritis, where they may be used for lubricating

CC mammalian joints, such as articulating joints of humans, dogs or horses.
CC The triboectectin, when formulated as a membrane, foam, gel or fibre, is
CC useful for inhibiting adhesion between two surfaces such as the injured
CC tissues of a mammal, where the injury is caused by a surgical insertion
CC or trauma, or an artificial device e.g., an orthopaedic implant. In
CC particular, one of the surfaces is pericardial tissue. DNA encoding a
CC triboectectin may be used in gene therapy. The present sequence represents
CC human MSF.
XX
SQ Sequence 1404 AA;

Query Match	100.0%;	Score 7410;	DB 22;	Length 1404;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1380;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	QULSSCAHRCGGVSRDQTCNDYQCMQHYMECCDPFKVCTAJELSCRCRCSFERRGEC	6
Db	25	qdlscagrcgegyaridatcmodycmqijymecoprfkrvtvlaelsckgrctesfgrtec	84
QY	61	DDDACCKYDYDCPPYEEFQCEVHNHPMPSPSKKAPRPSGASQITKSTTKSPKPNKK	120
Db	85	dcdagckkydrcopdyefcaeelmprpspkkarpipsagqtlstktsrpkpnkk	144
QY	121	TKKVISEETIEHSVSENOSSSSSSSSSSSTIMWKJSSKSNAANELQKTLVKONK	180
Db	145	lckviaseeelielchsvsengssssssssstlwlkssknaamelqlkllkvdnk	204
QY	181	KNRFRKKTTPRPVYDEGSLGNDQFVYTPRDNSTQHNKVSTSPKTTAKPKNRPSL	240
Db	205	knrtkkktrpkprvydeagsldngdlvtvtrpdsctlnhkvsspspklttkpnprrsl	264
QY	241	PNSDTSKEISLYVNKEETVETKEETTTNNQOSTDQKEKTSIAKETOSIEKTSAKDLAPT	300
Db	265	pnsdtsketa1vtnketvelctctlnqctsdgkxetsaketsioktsakdlapt	324
QY	301	SKVLAKPRPKAETTKGALPTKEBPPTPKRBASTPKRPPPTTKSAPTPKEBAPT	360
Db	325	skvlakprpkacttkgpaltrpkprptckrbastrckrppptctlkssaptrckaprt	384
QY	361	TTKSAPTPPKRBAPTTKKEBAPTTPKEBAPTTPKEBAPTTKSAPTPKRBAPTPKKA	420
Db	385	ttksaptrpkraprtctckepaptrckraprtctkraprtctkssaptrckraptrckpa	444
QY	421	PTTKRBAPTTPKRPPTTPPKRBAPTTPKRBAPTPKRBAPTPPKRBAPTTPK	480
Db	445	pttrckraptrckraptrckraprtckepaprtckraptrckraparckraprtckraptrck	504
QY	481	EPBAPTTKEBSPTTKRBAPTTKSAPTTKEBAPTTKSAPTTPKKEBSPPTTKRBAPTT	540
Db	505	epaprtckesprtrckraprtctkssaprtckepaprtckraptrckraprtckckepaprtc	564
QY	541	PKRBAPTPKKRAPTTTPKRBAPTTPKRBAPTTTKKRAPKRBAPTPPKRBAPTPPKL	600
Db	565	pkraptrckraprtckraprtckraprtckkpkraptrckraprtckraptrckraptrckkl	624
QY	601	TPPTPEKLABTPRKBAPTTPEELAPTTPEBPPTTPRKBAPTTPKAAANTPEKRAPTT	660
Db	625	tptrpkelaptrckraprtckraprtckraprtckraprtckraprtckraprtckraprtc	684
QY	661	PKRBAPTPKRBAPTTPKKEAPTTPKGTATTLKRBAPTTPKKPAKELAPTTKKEBPTST	720
Db	685	pkraptrckraprtckraprtckraptrckraprtctckraprtckrapkrapelaprtctkspstc	744
QY	721	TSDKRAPTTKGAAPTTPKRBAPTTPKGAAPTTPKGAAPTLLEKAPATTPKKPAKELAP	780
Db	745	tsdkraptrckraprtckraprtckraprtckraprtckraprtctkraprtckrapkrapelap	804
QY	781	TTTTKGPRSTTSOKRAPPTPKRBAPTTPKRBAPTTPKKRAPPTPEPRPTTSEVSTPPTTK	840
Db	805	cttkgprsttsdkraprtckraprtckraprtckraprtckkraprtckraprtckraprtckraprtc	864
QY	841	EPPTIHKSPDESTRPELSAETPRKALENSPKRBPVPTTKTPPAKPRMTTJAKTKTERPL	900

QY	1261	PPAATVGVGEMQVRRRREFRAAGSCQTPIRVSFPALAVADKCVLNEKVSLIMR	1320
Db	1285	rpaIrvpygemtqvrtrrrteraIgpseqtutitriqspalIaqdgylmnekvslIwv	1344
QY	1321	GLPVPWTSALSIPNIRRPDGYDYVAFSKDQYYNIDVPSRTARAITTRSGOTLSKWYNCP	1380
Db	1345	gIpnvvsalsIpnlrtpdgydyafakdqyInIdvpstraraittrsqtIskvwyncp	1404
RESULT	2		
ID	AAB60568	standard; Protein: 1404 AA.	
XX	AAB60568:		
XX	DT	27-APR-2001 (first entry)	
XX	DE	Human megakaryocyte stimulating factor (MSF, CACP).	
XX	DD		
KM	Human: CACP protein; campodactilyl-arthropathy-coxa vara-pericarditis;		
KM	MSF: megakaryocyte stimulating factor; synovial lubricant;		
KM	chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;		
KX	antiarthritic.		
XX	Homo sapiens.		
OS	WO200107068-A1.		
PN	01-FEB-2001.		
PD			
PF	21-JUL-2000; 2000WO-US20002.		
XX			
XX	23-JUL-1999; 99US-0145328.		
PR	19-JUL-2000; 2000US-0145328.		
XX	(UYCA-) UNIV CASE WESTERN RESERVE.		
PAA	Warman ML;		
PI	WPI; 2001-182721/18.		
DR			
XX	New composition comprising the campodactilyl-arthropathy-coxa		
PT	vara-pericarditis protein in combination with an anesthetic, useful for		
PT	treating osteoarthritis, or as lubricants of tissue and joints		
XX	Example 1; Page -: 34pp; English.		
XX	The invention relates to a method of treating osteoarthritis via the		
XX	administration of a composition comprising the campodactilyl-arthropathy-		
CC	coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.		
CC	The composition may further comprise a local anaesthetic. The composition		
CC	of the invention may be administered via intra-articular or intravenous		
CC	injection. The human CACP protein is identified in the invention as		
CC	being megakaryocyte stimulating factor (MSF). The gene encoding		
CC	CACP protein (MSF) is located on chromosome 1q25-31, and mutations in		
CC	this gene are responsible for the heritable disorder campodactilyl-		
CC	arthropathy-coxa vara-pericarditis, in which patients have synovial		
CC	hyperplasia without evidence of inflammation. CACP protein (MSF)		
CC	acts as a synovium lubricant, and can be used to lubricate tissue and		
CC	joints in the treatment of osteoarthritis. The composition may be		
CC	applied to reduce the symptoms of osteoarthritis (e.g., joint pain,		
CC	loss of range of movement or joint damage). The present sequence		
CC	represents human megakaryocyte stimulating factor (MSF, CACP protein).		
CC	Note: This sequence is not given in its entirety in figure 4 of the		
CC	specification, although a Genbank accession number was given. This		
CC	sequence was therefore obtained from Genbank (U70316).		
XX			
SQ	Sequence 1404 AA:		
Query Match	100.0%;	Score 7410;	DB 22; Length 1404;
Best Local Similarity	100.0%;	Pred. No. 0;	
Matches 1380; Conservative	0;	Mismatches	0; Indels 0; Gaps 0;

Qy	1	QDLSGCGRCGEGSYSDATONCDYMCQHYMECCDFEKRVCVTAELSCGKGFSEFEGRC	60
Db	25	qdlssacgrcgegyrdatcmcdyocqlymcecdpdkfrvctaelscqkcyfiesfergrec	84
Qy	61	DCDAQCKKYYKCCPRVDESCFAVHNPTSPSSKKAAPPSSGSGQIKSTTRSRSPKPNKK	120
Db	85	dcdagckkykccpdyesfcaevhnptspsskkaappsgsgqikstlrspkpnkk	144
Qy	121	TKKVISEETFEHSHVSENOESSSSSSSSSTLWIKIKSSKNSAANRELQKKLYKDKN	180
Db	145	tkkviaseetfeeshvseenoessssssssstlwikssksnaanrelqkklkydkn	204
Qy	181	KNRKKRPPTKPRVWDAGSGLDNGDFKVTYTPDSTTOHKNVYSPKITTAKFINRPSL	240
Db	205	knrkkrpkrpvywdaagsldngdfkvtctpdsctqhmvsybpklitakfinrpel	264
Qy	241	PNPNDTSKETSLVYNKETVWEKETTNNKQSTDGKSKTTSAKETOSIEKTSAKDLAPT	300
Db	265	pnndtsketeslvynketvwekettntnnkqstdgketsktsaketosiektsakdlapt	324
Qy	301	SKVIAPKPPKPAEITTTGCPALITTPKEPTPTTPKEBASSTPKPEPTTITKSAPTPKBPAP	360
Db	325	skviakppkpaetttgcpalittpkeptpttpkebasstpkpepttitskapitpkbpap	384
Qy	361	TTKSAPTTPKEBAPTTTKEBAPTPKEBAPTPKBPATTTKSAPTTPKEBAPTPKKA	420
Db	385	ttksapitpkebpittkebaptpkebaptpkbpatttksapitpkebpittkka	444
Qy	421	PTTPKEBAPTPKPEPTPTTPKEBAPTTTKEBAPTTTKEBAPPAKPKBAPTPPKBAPPTPK	480
Db	445	pttpkebapitpkpaptitpkebpapttkebpapttkebpaplkakpkapitpkebpaptk	504
Qy	481	BPAPTTPKEBPTTPKEBAPTTTKSAPTTPKBPATTTKSAPTTPKEBSPTTTKEBAPT	540
Db	505	ebpapttkebpittpkebpatttksapittkebpatttkebpatttkebpapttkebpapt	564
Qy	541	PKBPAPTPPKKBPATTPKKEBAPTTTPKKEBAPTTTKKBPAPAKKEBAPTTPKETAATTPKKL	600
Db	565	pkbpapitpkkbpattitpkebpattitpkebpattitkbpapakebapitpketapitkkl	624
Qy	601	TTTTPEKTLAPTPKBPAPTPPEELAPTPPEEPTTPPEEBAPTTTKKAAPTPKBPAPTT	660
Db	625	tttpeklapitpkbpaptppeelaptppeepttppeebaptttkkaaptpkbpapit	684
Qy	661	PKBPAPTPPKBPAPTPPKETAATTPPKGTAPTLIKBPAPTTPKKBPAPKELATTTKERTST	720
Db	685	pkbpapitpkbpapitpketapitpkgtapitlikbpapitpkkbpapkelatitkcertst	744
Qy	721	TPSDKAPTPPKGTAPTPPKBPAPTPPEBAPTPPKGTATTTKEBAPTPPKKBPAPKELAP	780
Db	745	tsdkapitpkgtaptpkbpaptppebaptpkgtatttkebpaptpkkbpapkelap	804
Qy	781	TTTTKPTSTSDKAPTPPKETAATTPKKEBAPTPPKKAPATTPETPPTTSEVSTPTTK	840
Db	805	tttkgptstsdkapitpketapitpkkebapitpkkapattpetpptsesvstpttk	864
Qy	841	BEPTTHKSPDSSTPLSAEPKPAJLNSPKPEPVPTTKPATKREBMTTKAKDKITERDL	900
Db	865	epitlhkspdsstpelseapklajlnspkpepvpttkpatkrebmttpakdkiterdl	924
Qy	901	RPTPEPTTAAPKMKETATTTTEKTTESKITATTTQVSTTTQDTPPTPKITTLKTTTAPK	960
Db	925	rttpepttaapmketatittekttteskitattttqvstttqodtpptpkittllkttlapk	984
Qy	961	VTTTKTITTEIANKPEETAKPKDKATNSKATTPPKOKPNTAKPKPSTGKPKMPVPR	1020
Db	985	vtttkitlitteimkpeetakpkdkatnskattppkokpntakpkpstgkpkmpvpr	1044
Qy	1021	KKPTPTPKKMSJMPBELNPTNSIAEAMLOTTTRPMQOTPSKILVENPKSEBAGABGET	1080
Db	1045	kptptpkkmsjmbelnptnsiaeamlotttrpmqotpskilevnpksebagabget	1104

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:15:40 ; Search time 107.17 Seconds
(without alignments)

Title: US-09-556-246-1_COPY_25_1404

Perfect score: 7410
Sequence: 1 QDLSSCAGRCGEGYSRDATC.....ARAITRSGQTLSKWNCP 1380

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
```

Database :

```
1: /SID58/gcgdata/geneseq/geneseqp/AA1980.DAT.*
2: /SID58/gcgdata/geneseq/geneseqp/AA1981.DAT.*
3: /SID58/gcgdata/geneseq/geneseqp/AA1982.DAT.*
4: /SID58/gcgdata/geneseq/geneseqp/AA1983.DAT.*
5: /SID58/gcgdata/geneseq/geneseqp/AA1984.DAT.*
6: /SID58/gcgdata/geneseq/geneseqp/AA1985.DAT.*
7: /SID58/gcgdata/geneseq/geneseqp/AA1986.DAT.*
8: /SID58/gcgdata/geneseq/geneseqp/AA1987.DAT.*
9: /SID58/gcgdata/geneseq/geneseqp/AA1988.DAT.*
10: /SID58/gcgdata/geneseq/geneseqp/AA1989.DAT.*
11: /SID58/gcgdata/geneseq/geneseqp/AA1990.DAT.*
12: /SID58/gcgdata/geneseq/geneseqp/AA1991.DAT.*
13: /SID58/gcgdata/geneseq/geneseqp/AA1992.DAT.*
14: /SID58/gcgdata/geneseq/geneseqp/AA1993.DAT.*
15: /SID58/gcgdata/geneseq/geneseqp/AA1994.DAT.*
16: /SID58/gcgdata/geneseq/geneseqp/AA1995.DAT.*
17: /SID58/gcgdata/geneseq/geneseqp/AA1996.DAT.*
18: /SID58/gcgdata/geneseq/geneseqp/AA1997.DAT.*
19: /SID58/gcgdata/geneseq/geneseqp/AA1998.DAT.*
20: /SID58/gcgdata/geneseq/geneseqp/AA1999.DAT.*
21: /SID58/gcgdata/geneseq/geneseqp/AA2000.DAT.*
22: /SID58/gcgdata/geneseq/geneseqp/AA2001.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	7410	100.0	1404	13	AAR26049	MSF precursor . Sy
2	7410	100.0	1404	22	AAB60568	Human megakaryocy
3	7410	100.0	1404	22	AAB29773	Human megakaryocy
4	6834	92.2	1299	22	AAB294322	Human bSF encoded
5	3484	47.0	902	22	AAB29778	Human MSF-derived
6	2081	28.1	452	16	AAB80041	Human megakaryocy
7	1707.5	23.0	472	22	AAB60569	Bovine MSF ortholo
8	1188	16.0	5179	22	AAB24516	C899P predicted am
9	981	13.2	763	21	AAC38942	Arabidopsis thaliana
10	950	12.8	1664	19	AAW43106	C. thermophilum O
11	805	10.9	1325	22	AAW03645	Peptide #2327 enco

12	152	10.1	182	12	AA10872
13	715	9.6	763	18	AA131852
14	645.5	8.7	4412	21	AA153666
15	613	8.3	111	13	AA26050
16	612	8.3	572	18	AA131855
17	555.5	7.5	804	7	AA60570
18	542	7.3	807	21	AA554617
19	520.5	7.0	788	21	AA55466
20	506.5	6.9	1837	21	AA11726
21	506.5	6.8	1474	9	AA62975
22	498.5	6.7	1721	21	AA11727
23	498.5	6.7	1721	19	AA448299
24	496.5	6.7	2971	21	AA11231
25	489.5	6.6	2972	22	AA50363
26	489.5	6.6	3118	22	AA50362
27	488.5	6.6	826	16	AA66042
28	488.5	6.6	816	13	AA16458
29	488	6.6	917	22	AA14187
30	488	6.6	957	21	AA592888
31	488	6.6	957	22	AA14513
32	485	6.5	1127	22	AA85541
33	476.5	6.4	652	9	AA62974
34	467.5	6.3	511	22	AA44883
35	467.5	6.3	511	22	AA17312
36	467.5	6.3	511	22	AA17312
37	454.5	6.1	1012	20	AA17406
38	454.5	6.1	2819	22	AA835608
39	450.5	6.1	378	12	AA161605
40	448	6.0	1237	21	AA14149
41	447.5	6.0	3266	21	AA142691
42	446.5	6.0	378	12	AA14149
43	446.5	6.0	750	20	AA105477
44	446.5	6.0	2665	22	AA14533
45	446.5	6.0	2665	22	AA126955

ALIGNMENTS

RESULT	1
AA26049	
ID	AA26049 standard; Protein: 1404 AA.
XX	
AC	AA26049;
XX	
DT	02-FEB-1993 (first entry)
XX	
DE	MSF precursor.
XX	
XX	Megakaryocyte colony stimulating factor; secretion signal; meg-CSF
KW	stability; proteolytic cleavage; adhesion; alternative splicing.
RM	

FH	Key	Location/Qualifiers
FT	Region	/label= Exon_I 1...26
FT	Region	/label= Exon_II 26..67
FT	Region	/label= Exon_III 67..107
FT	Region	/label= Exon_IV 107..157
FT	Region	/label= Exon_V 157..200
FT	Region	/label= Exon_VI 200..1141
FT	Region	/label= Exon_VII 1141..1166
FT	Region	/label= Exon_VIII 1166..1212
FT	Region	/label= Exon_VIII 1213..1266
FT	Region	/label= Exon_IX

Sequence of the fa
Mycobacterium tube
MSR-K130. Synthet
Sequence g1/01742
Mycobacterium tube
Protein encoded by
Human ORF22
Amino acid sequenc
Cytosporidium pa
Bioshensive precu
Portion of Cypriot
Cytosporidium pa
Human ORFX ORF995
Human SRCP. Homo
Human SRCP. Homo
P. yoelii SS2 ant
Peptide #2892 enco
Peptide #2869 enco
Human MUC11 polype
C90P predicted am
Human protein sequ
Bioshensive precu
Peptide #1317 enco
Peptide #1349 enco
Peptide #1289 enco
Human atropicin-1 f
Human ORFG57 gene
PRP 378. Trillicum
Streptococcus pneu
Human ORFX ORF225
PRP encoded by cld
C. albicans Rb1 f
Peptide #967 enco
Peptide #987 enco

1-3


```

XX 11-SEP-1984; 84AU-0007067.
PR 11-SEP-1984; 84AU-0007066.
PR 10-SEP-1985; 85AU-0047326.
XX
PA (HALL-) HALL INST MED RES.
XX
PI Kemp DJ, Anders R, Coppel RL, Brown G, Saint RB, Cowman AF;
DR MPI: 1986-094065/14.
DR N-PSDB: AAN60473.
XX
PT DNA coding for Plasmodium falciparum antigens - expressing
PT polypeptide(s) having antigenicity of RESA or FIRA antigens of P
XX falciparum
XX
PS Disclosure: Fig 7; 55pp; English.
XX
CC The inventors claim a novel DNA molecule which comprises a
CC nucleotide sequence corresp. to all or a portion of the base
CC sequence coding RESA (AAN60472) or FIRA (AAN60473). RESA and FIRA have
CC antigenicity suitable for providing protective immunity against
CC Plasmodium falciparum malarial infections.
XX
SQ Sequence 844 AA:

```

```

Query Match 7.6%; Score 555.5; DB 7; Length 844;
Best Local Similarity 25.7%; Pred. No. 2.8e-22;
Matches 252; Conservative 128; Mismatches 358; Indels 241; Gaps 49;

```

```

OY 111 ISEETIEHSVSENOESS-SSSSSSSSSTIMKIKSSKSANRELQKKLVKDNKN 168
Db 1 mesgkkaaspsinvdeyslsnneapnqatlnltpdqslahnlneinle----- 55
OY 169 RTKKRTPRPVVDAGSGLDNGDFKVTTPDSTIOHNVKSTSPKITAKPINRPSLP 228
Db 56 -----tlstltltnenevnpivpsis-ncldtclp 86
OY 229 NS-----DTSKEMSLYVKNKETVETKETTNNKOTSTGKEKTTSAKETQSIEKTS 279
Db 87 YQLIIIIISIVfckkkqkqlynk--nteeckfnlasvgsnatqgensnqknelev-kes 143
OY 280 AKDLAPTISKVLAKPTPKAETTTKGPALTTPKKEPT-----PTTPKKEPASTTPKEPTTIKS 335
Db 144 sqtqpv-----tpqetvtlq--aaatpgetvteqevltleevlt--qevlt--q 189
OY 336 APTTPKKEPATTTSAPTPPKKEPATTTPKKEPA-----PTTPKKEPATTTSAP 391
Db 190 epltvqep--vltvqevltvqep--pvtvqep--pvtvqepvltvqevltvqep--vtsqep 239
OY 392 TTPKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPA-----PTTPKKEPAPT 447
Db 240 vltvqep-----pvtvqep--pvtvqep--pvtvqepvltvqevltleevlt 284
OY 448 APKRPAPTTPKKEPAPTTPKKEPA-----PTTPKKEPAPTTPKKEPAPTTPKKEP 499
Db 285 -----qevltleev--pvtvqepvltleevltvqevltvqevltvqep--vltvqevltvqep 335
OY 500 APT-----TKSAPTTPKKEPAPTTPKKEPA-----PTTPKKEPAPT 531
Db 336 vtveehidekkqsegdnlsissiseeekshkckksswlfgrgnkndkksknekkps 395
OY 532 PTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTP 591
Db 396 leavkqnadeqkqpsdsgisvnaqsvltqepvtlqep-pltqelta----- 442
OY 592 EKLAAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEP 647
Db 443 -----ltq-----pvtvqepvltvqepvltvqepvltvqepvltvqepvlt 485
OY 648 KE-----PAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTP 703
Db -----pvtvqepvltvqepvltvqepvltvqepvltvqepvltvqepvltvqepvlt 703

```

```

Db 486 qeliatqepstqgehadekkasegdnlsisrlseeteekshkckksswlfgrgnkndk 545
OY 704 TSTSDKRAPPTPKGTAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEP 763
Db 546 ksknekkpslesvknadeqkqpsdsgisvnaqsvltqepvtlqep-pltqeltaq 604
OY 764 LAPTTTKGPTSTSDKRAPPTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEP 823
Db 605 -pltvq-----elvtqepvltvqepvltvqepvltvqepvltvqepvltvqepvltvqepvlt 648
OY 824 TTKKEPTTIHKSPEDE-----STPELSAEPPTKA-----LENS 854
Db 649 atqepstqgehadekkasegdnlsisrlseeteekshkckksswlfgrgnkndkkskn 708
OY 855 PKEPGVPTTKTPAATKPEMTTAKDCTTERDLRTTPE-TTAAAPKMTKEATTTTEKTTES 913
Db 709 ekpslesvknadeqkqpsdsgisvnaqsvltqepvtlqep-pltqeltaq 768
OY 914 KITATTTQVSTTTQODTTTFR--ITTLKTITTLAKKVTT-----TKKTTTTEIMNKPEET 966
Db 769 e-----vltqepvltvqepvltvqepvltvqepvltvqepvltvqepvltvqepvltvqepvlt 823
OY 967 AKPRDRATNSKATTPPKPOK 985
Db 824 asegdnlsisrlseeteek 842

```

```

Search completed: April 26, 2002, 16:25:13
Job time: 503 sec

```



```

QY 348 TKSAPTTKEPAATTTKEPAATTTKEPA---PTTTKEPAATTTKSAPTKEPAATTTK 403
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 805 eptpselptpselptpselptpselptpselptpselptpselptpselptpselpt 864
QY 404 KRAPTTKRPAPTTKRPAPTTKRPAPTTKRPAPTTKRPAPTTKRPAPTTKRPAPTTK 462
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 865 sdeplpsdeplps--deplps--deplps-sdeplpselptpselptpselptpselpt 919
QY 463 TTPKEPA--PTTTKEPSPT--TKEPAATTTKS--APTTPKEPAATTT--TTKSAPTTPKES 515
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 920 sdeplpsdeplpsdeplpselptpselptpselptpselptpselptpselptpselpt 979
QY 516 ----PTTTKEPAATTT--TKEPAATTTKRPAPTTKRPAPTTKRPAPTTKRPAPTTK 570
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 980 psdeplpsdeplpselptpselptpselptpselptpselptpselptpselptpselpt 1039
QY 571 APT---TTPKEPAATTTKRPAPTTKRPAPTTKRPAPTT---TPELAPTTTPEEPPT--TP 623
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1040 tpsdeplpselptpselptpselptpselptpselptpselptpselptpselptpselpt 1094
QY 624 EEPAPTTKAAAPNTPKBPAPTTKPEPA--PTTKEPAATTT--TPEKTAATTTKGTAPTTLK 680
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1095 eep1ptclpsdeplpsdeplpsdeplpsdeplpsdeplpselptpselptpselptpsd 1154
QY 681 BPAPTTKRPAPKELAPTTTKEPT--STSDKAPATTTKGTAPTTTKEPAATTTKRPAPTTT 739
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1155 eptps--deplpsd--eptpsdeplpselptpselptpselptpselptpselptpsd--deplps- 1208
QY 740 PKGTAPTTLKBPAPT--TPKRPAPKELAPTTTKGPTSTSDKAPATTT--TPEKTAATTTKE 795
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1209 ---deplpsdeplpselptpselptpselptpselptpselptpselptpselptpsd 1253
QY 796 PAPTTKRPAPAT--TPETPTPTTSEVSTPTTKEPTTHKSDESTPELSAPPTKALENS 854
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1254 ptps--deplpselptpselptpselptpselptpselptpselptpselptpselptpsd 1305
QY 855 PKPEGVPTTKTPAPAT--KPEMTTAKOKTTERDLRTTETTTAPAKMTKEATTTETKTE 912
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1306 sdeplpsdeplpselptpselptpselptpselptpselptpselptpselptpselptps 1361
QY 913 SKTATTTQVSTTTQDTPPKITTLKTTTLAPKVTTTKKITTTEIMNKPEETAKPKDR 972
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1362 deplpsdeplpselptpselptpselptpselptpselptpselptpselptpselptpsd 1396
QY 973 ATNSKAT-----TKRPOKPTKAP--KKPTSTKKKPTMPRVKRPKTTTPPKM----- 1017
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1397 sgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgs 1453
QY 1018 -----TSTMPELNPSTSRILAEAMLQTTTRPNQTSKLVENVNPKSDAGAGETPHM- 1069
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1454 yllgydgsgfpernltraeavlf-----akll-----gadesygaqgaaspsd 1498
QY 1070 LLRHH-----VMPPE-----VTPMDVLPVRYNNGI----- 1095
Db 1499 ladhwaawalkfatsgglfkgydgglfkpognltraeafatlvlnhfltkvkgelmskla 1558
QY 1096 ---TIINPMLSDETNICNG---KPVDLGLTLT-----RNGTIVAFRHHYMWMLSPFPPS 1442
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1559 tldslnpkfd---cvgnhagetlekltslgyisgydglt-----tkpqn 1600
QY 1143 PARITVWGIPIPIIDTVTRCNCEGKTFEKKD--SQVMFTNDIKD 1187
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1601 ylkrsesv---allnralergrlmgpklfpvnesywf-gdlnd 1642

```

RESULT 11

AA003645 standard: Protein; 1325 AA.

AC AA003645;

DT 09-OCT-2001 (first entry)

XX

```

DE Peptide #2327 encoded by probe for measuring breast gene expression.
XX Probe: human; breast disease; breast cancer; development disorder.
KW Inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX Homo sapiens.
XX MO200157270-A2.
XX 09-AUG-2001.
XX 29-JAN-2001: 2001WO-US00661.
XX 04-FEB-2000: 2000US-0180312.
XX 26-MAY-2000: 2000US-0207456.
XX 30-JUN-2000: 2000US-0608408.
XX 03-AUG-2000: 2000US-0632366.
XX 21-SEP-2000: 2000US-0234687.
XX 27-SEP-2000: 2000US-0236359.
XX 04-OCT-2000: 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR:
XX WPL: 2001-476286/51.
XX Novel single exon nucleic acid probe used to measuring gene expression
XX in a human breast.
XX Claim 27: SEQ ID NO 12385; 322pp; English.
XX
CC The present invention relates to novel single exon nucleic acid probes
CC (see AA100010-AA110067). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for measuring human gene expression in
CC a human breast sample, where the probe hybridises at high stringency to a
CC nucleic acid expressed in the human breast. The probes are useful for
CC predicting, diagnosing, grading, staging, monitoring and prognosing
CC diseases of the human breast, particularly those diseases with polygenic
CC aetiology. The diseases include: breast cancer, disorders of development,
CC inflammatory diseases of the breast, fibrocystic changes, proliferative
CC breast disease and non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1325 AA:

```

Query Match 11.0%; Score 802; DB 22; Length 1325;

Best Local Similarity 26.2%; Pred. No. 2.5e-35;

Matches 314; Conservative 116; Mismatches 506; Indels 264; Gaps 46;

```

QY 61 KRYCTAAEYHNPTSPSSSKKAPPGASQTI--KSTTKSPKPP-----NKKRTKKVI 111
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 56 krcncltrnshkprd-----kptgnskltdhkselchheaprtseenssqgxdmrlr 107
QY 112 ESEELTEHVSSENOESSSSS-----SSSSSTTITWKIKSKNSAANRELOK 158
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 108 nqrsvpadstthkkesagkhilpapakincrstsktstv-----tkskdktgplek 163
QY 159 KILKVP-----NKKNRTKKKPPKPPVVDAGSGLDNGDFKVT--TPDTST 202
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 164 smstldktsashklttsfhmgnsqtkqsktsfpekltaa-----sktykctgtpese 219
QY 203 TOHNKYST--SPKITTAKEINRPSPSLPPV-----SDTSKESLIVNKKETTYETKETT 254
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 220 kredsrtlvaskliltktknigetlsanelqtslaaepelnggrtanennppspapten 279
QY 255 NKQSTGDEKETT--SAKETQSIKTSANDLAP-----TSKVLAKPPTPK 296
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 280 rerta---nentllspaeptenrentanentapfapagptenentanentllfapapten 336

```

```

PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 13.4%; Score 981; DB 21; Length 763;

Best Local Similarity 38.5%; Pred. No. 3.3e-45; Matches 226; Conservative 24; Mismatches 293; Indels 44; Gaps 11;

```

OY 298 ETTTKGALTTPEKPTTKEPASTTPEKPTTTSAPTTEKPAATTTKSAPTTKE 357
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 126 qtdagunpylirppriipppcvcicp-epsf-----ppptvt---pqippt 170
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 358 PAPTTTKEPAATTTKEPAATTTKEPAATTT--TKSAPTTKEPAATTTKKEPA 415
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 171 lptlptpetppptppapreivlpptlppqptlppqptlppqptlppqptlpp 226
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 416 TTPKEPTTTPKEPAATTT--KEPAATTTKEPAATTTKKEPAATTTKEPAATTT 474
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 227 ntpetpcttppetppetppetppetppetppetppetppetppetppetpp 286
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 475 EPSPTTKEPAATTTKSAPTTKEPAATTTKSAPTTKEPESTTTPKEPAATTTKE 534
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 287 etpntpctpsctppetppetppetppetppetppetppetppetppetppet 340
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 535 PKKPAATTTKEPAATTTKEPAATTTKKEPAATTTKEPAATTTKEPAATTTKE 594
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 341 ppgtppntppetppetppetppetppetppetppetppetppetppetppet 400
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 595 APPTPEKPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTT 654
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 401 lptnppgtppetppetppetppetppetppetppetppetppetppetppet 460
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 655 PKKPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKE 714
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 461 pklp-ppctppetppetppetppetppetppetppetppetppetppetpp 514
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 715 TPKGTAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKE 774
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 515 lptlptppqspqilppllpntppetppetppetppetppetppetppetpp 574
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 775 TTSDEP-----APPTPEKPAATTTKEPAATTTKEPAATTTKEPAATTTKE 827
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 575 lptnppgtppetppetppetppetppetppetppetppetppetppetppet 634
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 828 P---TTIHKSPDESTPELSAETPKALENSPKKPGVPTTKKPPAATTKP 871
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 635 pppttllppspstppnspsppppksqpppppprfrgpppppprctcp 681
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 10
AAW43106
ID AAW43106 standard; Protein; 1664 AA.
XX AC AAW43106;
XX 16-OCT-1998 (first entry)
XX C. thermocellum O1pB protein.
DE

```

XX OS Clostridium thermocellum.
XX FH Key
XX FH Location/Qualifiers
XX FT Domain
XX FT 28..192 "cohesin type II domain"
XX FT /note="cohesin type II domain"
XX FT 207..363
XX FT /note="cohesin type II domain"
XX FT 409..565
XX FT /note="cohesin type II domain"
XX FT 607..763
XX FT /note="cohesin type II domain"
XX PN FR2748479-A1.
XX PD 14-NOV-1997.
XX PE 10-MAY-1996; 96FR-0005854.
XX PF 10-MAY-1996; 96FR-0005854.
XX PR 10-MAY-1996; 96FR-0005854.
XX PA (INSP ) INST PASTEUR.
XX PI Beguin P, Leibovitz E;
XX PE WPI; 1998-011569/02.
XX DR N-PSDB; AAW86623.
XX PT Cellulase proteins with cohesin or dockerin type II domains - useful
XX PS for potentiating the activity of multiprotein enzyme complexes
XX PS Claim 7; Page 31-39; 60pp; French.
XX CC Multimeric protein, especially enzymatic, complexes are held together
XX CC by protein-protein interactions between domains designated dockerins
XX CC and cohesins, which are found on the catalytic and scaffold subunits
XX CC respectively. An example of such a complex is the cellulose degrading
XX CC protein complex from Clostridium thermocellum, known as the cellulosome.
XX CC This complex comprises around 15 proteins including endoglucanases,
XX CC cellobiohydrolases, hemicellulases, e.g. xylanases or lichenases, which
XX CC interact with a central "scaffold" protein designated the cellulosome
XX CC integrating protein (CipA; see AAW43108). The catalytic subunits
XX CC interact with the Cip subunit via conserved 23 amino acid dockerin
XX CC domains. Cip has been shown to contain 9 copies of a cohesin domain.
XX CC The invention relates to the isolation of proteins binding to a novel
XX CC dockerin type domain found in the C-terminal portion of Cip. The new
XX CC I domain found on the catalytic subunits of the cellulosome). The type
XX CC II dockerin domain has some sequence similarity to the type I dockerins
XX CC but is unable to bind type I cohesin domains.
XX CC The sequence presented here is an example of a protein which binds
XX CC the novel type II dockerin domain and is the product of the O1pB gene.
XX CC The protein contains 4 type II cohesin domains in the N-terminal portion
XX CC of which the first domain (amino acid residues 28-192) is thought to
XX CC bind CipA. The novel type II dockerin and cohesin domains can be used
XX CC in complexes, especially enzyme complexes, to potentiate their catalytic
XX CC actions in a synergistic manner.
XX SQ Sequence 1664 AA;

```

Query Match 13.0%; Score 950; DB 19; Length 1664;
Best Local Similarity 31.0%; Pred. No. 3.5e-43;
Matches 312; Conservative 104; Mismatches 361; Indels 230; Gaps 55;
OY 289 VLAKPTP-KAETTTKGALTTPEKPTTKEPASTTPEKPTTTSAPTTEKPAATTT 347
DB 758 vvlqppikaasdepilddtppsdepts-----depts-----deptsdeptsd 804

PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132865.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140891.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142877.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150566.
PR 27-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0151830.
PR 10-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 15-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154039.
PR 22-SEP-1999; 99US-0154779.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155659.
PR 29-SEP-1999; 99US-0156458.
PR 04-OCT-1999; 99US-0156596.
PR 05-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157753.
PR 07-OCT-1999; 99US-0157865.
PR 08-OCT-1999; 99US-0158029.
PR 12-OCT-1999; 99US-0158232.
PR 13-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.

CC and in assays to identify modulators of TCAP expression and activity.
 CC Anti-(1) antibodies and antagonists may also be used to down regulate
 CC TCAP expression and activity. The anti-(1) antibodies may also be used
 CC as diagnostic agents for detecting the presence of TCAPs in samples
 CC (e.g. by enzyme linked immunosorbent assay (ELISA)). AA1284460 to AA129512
 CC and AA4494 to AA424523 represent nucleotide and amino acid sequences
 CC given in the exemplification of the present invention.

XX Sequence 5179 AA;

Query Match 16.1%; Score 1171; DB 22; Length 5179;

Best Local Similarity 29.2%; Pred. No. 1.5e-54;
 Matches 399; Conservative 79; Mismatches 524; Indels 364; Gaps 52;

```

QY 29 SCAGRCGEGYSRDATCNCQYHMEC-----CPDKRKVCTAAEVHNPIS 74
  || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1079 scs--cdtg--gdcecfcsasayagcctkcgacvfwtrpdlcpif-----cdyyprph 1128
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 75 PPSKKAPPSGASQRTIKSTT-----KRSRK--PNNKKTKKVIESP-- 115
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1129 ecewhypcgnrfetortlnglnhsnsvsylegcyprcpxdrplyealdkcvtaekg 1188
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 116 --ITEEH--SVSENOSSSSSSSSSSSTI-----WKIKSSKNS 150
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1189 cyvedthypgaaavpteetkscvctnsqgvrcpreegkllnqdgdaefcyweI-----c 1243
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 151 AANRELQK-----KIKVKNKKNKTKKPKPPVY----- 181
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1244 gprgtvekhnicsltprstlftltltpitpftsftlttlttcsstvlsttpklcc 1303
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 182 -----DEAGSLDNGDK-----VTTPDSTTQH--NKVSTSPK--- 213
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1304 lwsdwinedhpsqsgddgdrpfidgvcgapedlectrsvkophslsqhngqkvqcdvsvf 1363
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 214 -----ITTAKPINDRPSLPPNSDTSKETSILTVN 241
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1364 icknedfgngpfglcydykrlrvncwpmdkcltpprttpprttltltpitps 1423
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 242 KETTVETKETTNNKQSTQCKEKTSAKETQSIKTSANDLATSVLAKPTPKAETTT 301
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1424 prlt-----tltppttspilttltt-----lpt-----tpspilstt 1462
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 302 KGPLTLTPKEPT-----PTPKEPASTTPKEPTTIKSAP--TPPKP-----APTTP 348
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1463 tpprttpprttsppttsppttlttpprttpprttpprttpprttpprttpprtt 1522
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 349 KSAP--TPPKEPAPTTTKEPAPTTKEP-----APTTPKEPAPTTKSAP--TPKEPA 398
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1523 psppttlttpprttsppttlttpprttlttpprttlttpprttlttpprttsppt 1582
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 399 PTTPKKAPPTPKEPAPTTTPKEPTTPPKEPAPTTKEPAPTTKEPAPTTKEP-----APTAPKK 451
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1583 tltpspilttlttpprttsppttlttpprttlttpprttlttpprttlttpprttsp 1642
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 452 PAPTTPKEPAPTTTPKEPAPTTTPKEPSPTTPKEPAPTTTKSAPTTPKEPAPTTKSAP--- 508
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1643 ppprttlttpprttsppttlttpprttlttpprttspilttpprttspilttpprt 1702
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 509 TTPKPSPTTTPKEPAPTTTPKEPAPTTTPKKAAPTTPKKAAPTTPKKAAPTTPK 568
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1703 tltpspilttlttpprttsppttlttpprttlttpprttlttpprttlttpprttsp 1761
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 569 EPAPTTP-----KETAPTTP----- 583
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1762 ppsltpprfspfsttlttppcplcnwtglwdsqkpnfhkpggdeltlqdcvgpwaani 1821
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 584 ---KKIAP----- 590
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1822 scratmpdvplgqlgvtvcdvsvglcknedkpgvgvlpmafclnyelnvqcecvtlg 1881
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 591 PEKLAPTTPPKAPPTPEELAPTTPEEPTPT--TPPEEPAPTTPKAAAPTPKEPAPTTTKE 649
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

Db 1882 plmtltttenprrtlttlttlttvtprpvtgltprlttlttlttvtprpvtgt 1941
QY 650 P-----APTTPKEPAPT--TPKETAPTTPKGTAPTTTKEPAPTTPKKAPKELAPTTKEP 703
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1942 ptttlttlttvtprpvtgltprlttlttlttvtprpvtgltprlttlttlttvtpr 2001
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 704 TSTTSDKRAP--TPKGTAPTTKEPAPTTTPKKAAPTTPKGTAPTTTKEPAPTTPKKAPK 762
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2002 tprpvtgltprlttlttlttvtprpvtgltprlttlttlttvtprpvtgltpr 2060
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 763 ELAPTTPKGTSTTSDKRAP--TPKETAPTTPKKAAPTTPKKAAPTTPKEPTTSEVST 821
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2061 pilttlttvtprpvtgltprlttlttlttvtprpvtgltprlttlttlttvtpr 2119
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 822 PTTTPKEPTTPKGTSTTSDKRAP--TPKETAPTTPKKAAPTTPKKAAPTTPKEPTTSEVST 870
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2120 prpvtgltprlttlttlttvtprpvtgltprlttlttlttvtprpvtgltpr 2176
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 871 PEMTTAKDKTTERDRL--TRP--ETTAAAPKMT--KETATTEKTESKITATTOY 922
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2177 tlttlttvtprpvtgltprlttlttlttvtprpvtgltprlttlttlttvtpr 2235
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 923 TSTTSDKRAP--TPKGTAPTTKEPAPTTTPKKAAPTTPKKAAPTTPKKAAPTTPK 981
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2236 tprpvtgltprlttlttlttvtprpvtgltprlttlttlttvtprpvtgltpr 2291
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 982 KPQKPTKAP--KRPSTPKKPTMPRVKRPKTPTPPKMTSTMPELNP---TSRIAEAMLQT 1037
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2292 tlttlttvtprpvtgltprlttlttlttvtprpvtgltprlttlttlttvtpr 2351
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1038 TTR--PNOGTPNSKLVENPKSEDAGAGETPHMLLRHVMPREYTP 1082
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2352 gtqprlttlttlttvtprpvtgltprlttlttlttvtpr 2392
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 9

AA638942 standard; Protein: 763 AA.

AA638942.

18-OCT-2000 (first entry)

Arabidopsis thaliana protein fragment SEQ ID NO: 48115.

Protein identification: signal transduction pathway; metabolic pathway;

hybridisation assay; genetic mapping; gene expression control; promoter;

termination sequence.

Arabidopsis thaliana.

EP1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-0301439.

25-FEB-1999; 99US-0121825.

05-MAR-1999; 99US-0123180.

09-MAR-1999; 99US-0123548.

23-MAR-1999; 99US-0125788.

25-MAR-1999; 99US-0126284.

29-MAR-1999; 99US-0126785.

01-APR-1999; 99US-0127462.

06-APR-1999; 99US-0128234.

08-APR-1999; 99US-0128714.

16-APR-1999; 99US-0129845.

19-APR-1999; 99US-0130077.

21-APR-1999; 99US-0130449.

23-APR-1999; 99US-0130510.

23-APR-1999; 99US-0130891.

28-APR-1999; 99US-0131449.

30-APR-1999; 99US-0132048.

```

XX AC AAB60569;
XX 27-APR-2001 (first entry)
XX DT
XX DE Bovine MSF orthologue, superficial zone protein (SZP).
XX KM Bovine; CACP protein; camptodactylly-arthropathy-coxa vara-pericarditis;
XX KM superficial zone protein; SZP; MSF orthologue; synovial lubricant;
XX KM osteoarthritis; joint lubrication; osteopathic; antiarthritic.
XX OS Bos taurus.
XX PN WO200107068-A1.
XX PD 01-FEB-2001.
XX PE 21-JUL-2000; 2000WO-US20002.
XX PR 23-JUL-1999; 99US-0145328.
XX PR 19-JUL-2000; 2000US-0145328.
XX PA (UYCA-) UNIV CASE WESTERN RESERVE.
XX P1 Warman ML;
XX DR WPI: 2001-182721/18.
XX PT New composition comprising the camptodactylly-arthropathy-coxa
XX PT vara-pericarditis protein in combination with an anesthetic, useful for
XX PT treating osteoarthritis, or as lubricants of tissue and joints
XX PS Example 1; Fig 4; 34pp; English.
XX CC The invention relates to a method of treating osteoarthritis via the
XX CC administration of a composition comprising the camptodactylly-arthropathy-
XX CC coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.
XX CC The composition may further comprise a local anesthetic. The composition
XX CC of the invention may be administered via intra-articular or intravenous
XX CC injection. The human CACP protein is identified in the invention as
XX CC being megakaryocyte stimulating factor (MSF). The gene encoding
XX CC CACP protein (MSF) is located on chromosome 1q25-31, and mutations in
XX CC this gene are responsible for the heritable disorder camptodactylly-
XX CC arthropathy-coxa vara-pericarditis, in which patients have synovial
XX CC hyperplasia without evidence of inflammation. CACP protein (MSF)
XX CC acts as a synovium lubricant, and can be used to lubricate tissue and
XX CC joints in the treatment of osteoarthritis. The composition may be
XX CC applied to reduce the symptoms of osteoarthritis (e.g., joint pain,
XX CC loss of range of movement or joint damage). The present sequence
XX CC represents the bovine orthologue of human MSF, superficial zone
XX CC protein (SZP).
XX SQ Sequence 472 AA;

```

```

Query Match 23.48; Score 1707.5; DB 22; Length 472;
Best Local Similarity 79.98; Pred. No. 1,7e-84;
Matches 319; Conservative 24; Mismatches 53; Indels 3; Gaps 3;

```

```

Db 254 tewwgsipidvtrcmcegtffkfgsgywrftndikdagypk1iskgfyglnqkiva 313
Qy 1208 ALSTAKYKMPESVYEFKRGSGIDQYIKQEVQKCPGRPALNPVYGEMQVRRRFE 1267
    ||| :||: ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 314 alslegysrpesvvyffkrgsgvgytkqepiqctisrpalnsvygelcaqvrrirfe 373
Qy 1268 RAIGSQTHTRIQXSPARLAYQDKGLHNEKYSILMRGLPNVNTSAISLPNIRKPDGY 1327
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 374 raigpsqvtlrlhytlpvrpygdkglhnekvksclwrglpnvvtalsipnlrkpdgy 433
Qy 1328 DYAFSKDQYNIIDVPSRATITRRSGOTLSKWNCP 1366
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 434 dyalskdqyynidvpsrtatitrrsgqlsnwncp 472

RESULT 8
ID AAM24516
ID AAM24516 standard; Protein: 5179 AA.
XX AC AAM24516;
XX DT 12-OCT-2001 (first entry)
XX DE C899P predicted amino acid sequence.
XX KM Human; immunotherapy; diagnosis; colon cancer; colon tumour;
XX KM immunogenic; gene therapy; vaccine; colonic cancer.
XX OS Homo sapiens.
XX PN WO200149716-A2.
XX PD 12-JUL-2001.
XX PE 29-DEC-2000; 2000WO-US35596.
XX PR 30-DEC-1999; 99US-0476296.
XX PR 10-JAN-2000; 2000US-0480321.
XX PR 15-FEB-2000; 2000US-0504629.
XX PR 06-MAR-2000; 2000US-0519444.
XX PR 19-MAY-2000; 2000US-0519251.
XX PR 29-JUN-2000; 2000US-0609448.
XX PR 28-AUG-2000; 2000US-0649811.
XX PA (CORI-) CORIXA CORP.
XX PI Xu J, Lodes MJ, Secretist H, Benson DR, Meagher MJ, Stolk JA;
XX PI King GE, Wang T, Jiang Y;
XX DR WPI: 2001-441847/47.
XX PT Colon tumor associated proteins and nucleic acids useful for the
XX PT prevention, diagnosis and treatment of colonic cancer.
XX PS Claim 2; Page 446-462; 472pp; English.
XX CC The present invention describes colon tumour associated proteins (I) and
XX CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
XX CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
XX CC (II) may be used in the prevention, diagnosis and treatment of diseases
XX CC associated with inappropriate colon tumour associated protein (TCAP).
XX CC expression, such as colonic cancer. For example, (I) and (II) may be
XX CC used to treat disorders associated with decreased expression by
XX CC rectifying mutations or deletions in a patient's genome that affect the
XX CC activity of TCAPs by expressing inactive proteins or to supplement the
XX CC patient's own production of them. Additionally, (II) may be used to
XX CC produce the TCAP proteins, by inserting the nucleic acids into a host
XX CC cell culturing the cell to express the protein. (II) and its
XX CC complementary sequences may also be used as DNA probes in diagnostic
XX CC polymerase chain reaction (PCR) and hybridisation assays to detect and
XX CC quantitate the presence of similar nucleic acids in samples, and
XX CC therefore which patients may be in need of restorative therapy. (I) may
XX CC also be used as antigens in the production of antibodies against TCAPs

```



```
Db 601 aptapkeapcttckaptltpcklcttpecklapttpekapcttpeelapttpeepctt 660
Oy 663 PEERAPTPKAAAPNTPKREPATTTPKREPATTTPKREPATTTPKREPATTTPKREP 682
Db 661 peepapctcpkaapnctpkrepapcttckrepapcttckrepapcttckrepapcttck 720
Oy 663 APPTPKKAPKELAPTTTKEPTSTSDKPAPTTPKGTAPTTTKEPAPTPKREPATTTPK 742
Db 721 apttpckpapkclapttckpstsdckpapttckpapttckpapttckpapttckp 780
Oy 743 TAPTTLKEPAPTPPKKAPKELAPTTTGTPTSTSDKPAPTTPKGTAPTTTPKREPATTTPK 802
Db 761 tapttckrepapcttckpapkclapttckpapttckpapttckpapttckpapttckp 840
Oy 803 KPAPTPEPTPEPTSEVTPPTTKEPTTHKSPDSEBELSAPTPKALENPKRPGVPT 862
Db 841 kpaPtPePtPePtSeVtPcttckpcttckpcttckpcttckpcttckpcttckpcttckp 900
Oy 863 TKTPAATKPEMTTAKDKTTERDLRTPEPTTAAPKMTKETATTEKTESKATATTQV 922
Db 901 tktpaatkpeamttaKdkTterdlrtPePtTaaPkMtKeTAtTeKtESKATATTQV 960
Oy 923 TSTTQDTTPPKRTTLTKTTTAAPKVTTTKTITTTTEIMNKPEETAKPKDRATNSKATTPK 982
Db 961 tSttqdtctPtKtltcttclapktcttckkltctclmKpeetAkPkDrAtNsKAtPtk 1020
Oy 983 POKPTAKKKPTSTPKPTMPRVKPKTTPPRKMTSMPELNPSTRIAEAMLOTTTSPN 1042
Db 1021 pOkPtAkKkPtStPkPtMpRvKPkTtPPrKMTsMpELnPsTrIAeAMlOtttSPn 1080
Oy 1043 QTPNSKLVNPKSEDAAGAGETPRHMLLRPHVPEVTPMDVLYPRVNOGIILNPMIS 1102
Db 1061 qTpNsKlVnPkSeDaAGeTPrHmLLrPhVPeVtPMDvLyPrVnOGiIlNpMiS 1140
Oy 1103 DETNCGNKRPVDSLTTLANGLVAFRGHYFWMLSPFSPSARRTTEWGISPDYFT 1162
Db 1141 deTnCGnKRPvDslTtLanGLvAFrGHYfWmLsPsfSPsARrtTEwGIsPdYft 1200
Oy 1163 RCNCEGKTFEPRKDSQYMRFTNDIKDAGYPRKPIFKGFGILTGOIYALSTAKYKMPESVY 1222
Db 1201 rcnCeGkTfEPrKdSQYmRfTnDiKdAGyPrKpIFkGfGILtGoIyAlStAkYkMpEsVY 1260
Oy 1223 FFKRGGSIQOITTYKQEPYQKCPGRPALNLYVGE 1257.
Db 1261 fFkRgGSiQOItyTyKQEPyQkCpGrPaLnLyVgE 1295

RESULT 5
AAB29778 standard; Protein; 902 AA.
AC AAB29778;
DT 28-FEB-2001 (first entry)
DE Human MSF-derived tribonectin.
KW Human tribonectin; MSF; megakaryocyte stimulating factor;
KW alternative splicing; joint boundary lubricant; O-linked oligosaccharide;
KW osteoarthritis; tribonecplementation; tissue adhesion inhibition;
KW friction coefficient reduction; gene therapy; antiarthritic;
KW osteopathic.
XX Homo sapiens.
XX OS
XX PN WO200064930-A2.
XX PD 02-NOV-2000.
XX PF 24-APR-2000; 2000WO-US10953.
XX PR 23-APR-1999; 99US-0298970.
```

```
XX
PA (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.
XX
PI Jay GD:
XX
DR WPI; 2001-024673/03.
XX
PT Novel tribonecclin polypeptide useful as lubricant for treating
PT osteoarthritis, comprises O-linked lubricating moiety
XX
PS Disclosure: Fig 1; 47pp; English.
XX
CC The invention relates to a human tribonecclin which is a product of
CC alternative splicing of the human MSF (megakaryocyte stimulating factor)
CC gene. The tribonecclin has at least one O-linked oligosaccharide
CC lubricating moiety and has a polypeptide sequence comprising 1-76
CC repeats of a motif having at least 50% identity to the sequence KEAPPT
CC (AAB29774). The invention also relates to a nucleic acid encoding a
CC human MSF-derived tribonecclin; a biocompatible composition comprising a
CC human tribonecclin for inhibiting tissue adhesion formation; and a method
CC of diagnosing osteoarthritis or a predisposition to osteoarthritis by
CC measuring the amount of MSF or its fragment in a biological sample of a
CC mammal, wherein an increased amount of MSF compared to a control
CC indicates the presence of or predisposition to developing
CC osteoarthritis. The tribonecclin and DNA encoding it are useful in the
CC treatment of osteoarthritis, where they may be used for lubricating
CC mammalian joints, such as articulating joints of humans, dogs or horses.
CC The tribonecclin, when formulated as a membrane, foam, gel or fibre, is
CC useful for inhibiting adhesion between two surfaces such as the injured
CC tissues of a mammal, where the injury is caused by a surgical insertion
CC or trauma, or an artificial device e.g., an orthopaedic implant. In
CC particular, one of the surfaces is pericardial tissue. DNA encoding a
CC tribonecclin may be used in gene therapy. The present sequence represents
CC a substantial portion of a human MSF-derived tribonecclin.
XX
SQ Sequence 902 AA;

Query Match 47.8%; Score 3484; DB 22; Length 902;
Best Local Similarity 74.6%; Pred. No. 9.6e-180;
Matches 794; Conservative 26; Mismatches 80; Indels 162; Gaps 66;

Oy 162 VKDNKKNTKKKPPKPPVNDAGSGLDNGDFKVTTPDSTYOHNNKYSTSPKITTAKPIN 221
Db 1 vkdnknrtkkppkppvndagsgldngdfkvttpdstyohnnkystspkittakin 60
Oy 222 PRPSLPNSDSKETSLSLVNKEETVETKETTNNKOSTDGEKETSAAKETSIEKTSAK 281
Db 61 prpslpsnstdsketslslvnketvetttnkqstgdgkelsaketsiektsak 120
Oy 282 DLAPTSKVIAPKPKAETTTGPAALTTTPKEPTTPPKREPASTTPKREPPTTIKSAPTPK 341
Db 121 dlaptskviakpPKAETTTGpAALtTTPKEPTTPPKREPASTTPKREPPTTIKSAPtPk 172
Oy 342 EPAPTTKSAPTTPKREPAPTTTPKREPAPTTTPKREPAPTTTPKSAPTTPKREPAPTT 401
Db 173 epapttksapttPKREPAPTTTPKREPAPTTTPKREPAPTTTPKSAPtTPKREPAPtT 228
Oy 402 PKRAPPTPKREPAPTTTPKREPAPTTTPKREPAPTTTPKREPAPTTAKKAPPTPKEP 461
Db 229 PKRAPPTPKREPATtTPKREPAPTTTPKREPAPTTTPKREPAPTTAKKAPtTPKEP 281
Oy 462 PTPPKREPAPTTTPKREPAPTTTPKREPAPTTTPKSAPTTTPKREPAPTTTPKEPSTTKE 521
Db 282 pttPKREPAPTTTPKREPAPTTTPKREPAPTTTPKSAPtTPKREPAPTTTPKEpSTtKe 335
Oy 522 PAPTPPKREPAPTTTPKREPAPTTTPKREPAPTTTPKREPAPTTTPKREPAPTTPKETA 581
Db 336 papttPKREPAPTTTPKREPAPTTTPKREPAPTTTPKREPAPTTTPKREPAPtTPKETA 388
Oy 582 TPKKLPTTPPEKLIAPTTPEKPAPTTPELIAPTTPEEPPTTPPEEPATTTKAAAPNPKPE 641
Db 389 t-PkEApTtPeKlIAPtTPEKPApTTPElIAPtTPEEPtTPPEEPATtTKAAAPNPKPe 440
```

DB 721 apitpkpapaklapitpketstsdskpaptkpglapitpkpaptkpg 780
 QY 743 TAPPTKEPAPTPPKKAPAPELAPTTKGTSTSDKAPATTPPKKAPATTPPK 802
 DB 781 tapitkpeaptpkpkpapeiapitkpgtstsdskpaptkpglapitpkpaptkpg 840
 QY 803 KPAPTPEPTPPPTSEVSTPTTKEPTTHKSDESTPELSAEPPTKALENSKEPVP 862
 DB 841 kpaptkpeaptpptsevtpttketphtkspdestpelsaepptkalenskepvp 900
 QY 863 TKTPPAKPEMTTAKDKTTERDLRTPEPTTAAPKMTKATATTEKTSKIRATTOY 922
 DB 901 tktpaakpemtakdkkterdlrtpepttaapkmktakattektskiralatcgv 960
 QY 923 TSTPTODTPEPKITTTKTTTLAPKVTTKTKTTTTEIANKPEETAKPKDRATNSKATTPK 982
 DB 961 tsitttdtptfkittlttclapkvtttkitttletimnkpsetakpkdratnskattpk 1020
 QY 983 POKPTAPKPKPTSTKPKTKMPRVAKPKTTPPKMTSTMEPLNPTSKIAEAMLOTTTRPN 1042
 DB 1021 pdkptkapkpkstckpkmpvrpkptktpkmtstmpelnpstriaeamlqtltrpn 1080
 QY 1043 QTPMSKIVENPKSEDSGAGEPTPHMLLRPHVFMPEVTPMDVLPVPNOGIINPMLS 1102
 DB 1081 qtpmskivevnpksedsgageetphmllrphvfmpevtpmdvlpvpnggiiinpmls 1140
 QY 1103 DETNICKGKSPVDGLTTLRLNGTLVAFRGHYFMWLSFSPSPSPARRITEVWGISPTDVT 1162
 DB 1141 detnckgkspvdglttlrngtlvafrghyfmwlsfspsparritlevwigspstidvft 1200
 QY 1163 RCNNEGTFEPKDSQVWRFTNDIDAGYPPKPKFGSLGSOIYAAISTAKYKMPESVY 1222
 DB 1201 rcnnegtfifkdsqwrftndidagypkpfkfglsglvaastakaykmpesvy 1260
 QY 1223 FFRKGSIOQYIYKQEPVOKCPGRPPALNYPVYGEMLQVRRRPERAIGPSQHTTIRIOY 1282
 DB 1261 ffrkgsgioqyiykqevpvcgrrppalnpvygemlqvrrrrferalpgsqhtlirioy 1320
 QY 1283 SPARLAVODKGVLEHNEKVSILMRGLDPNVVTSALSLPINKRPDGYDYAASKDOYINIDV 1342
 DB 1321 sparlaydkgvlehnkvslmrglpnyvtsalslpinkrpdgydyafaskdgyinidv 1380
 QY 1343 PSRTARAITTRSGOTLSKRWYNCP 1366
 DB 1381 psrtaraaitrsgqlskrwyncp 1404
 RESULT 4
 AAM24322
 ID AAM24322 standard; Protein: 1299 AA.
 AC AAM24322;
 XX 12-OCT-2001 (first entry)
 DT
 XX
 DE Human EST encoded protein SEQ ID NO: 1847.
 XX Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
 KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
 KW diagnostics; forensic test; gene mapping; genetic disorder;
 KW biodiversity; gene therapy; nutrition.
 OS Homo sapiens.
 XX
 PN WO200154477-A2.
 XX
 PD 02-AUG-2001.
 XX
 XX 25-JAN-2001; 2001WO-USO2687.
 PF
 XX 25-JAN-2000; 2000US-0491404.
 PR 17-JUL-2000; 2000US-0617746.
 PR

PR 03-AUG-2000; 2000US-0631451.
 PR 15-SEP-2000; 2000US-0663870.
 XX
 XX (HYSE-) HYSEO INC.
 PA
 XX Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
 PI Cao Y, Drmanac RA, Zhang J, Werlman T;
 XX
 XX WPI; 2001-4/76164/51.
 DR N-PSDB; AAH98981.
 XX
 PT Isolated polypeptide for treatment of diseases, diagnostics, raising
 PT antibodies and research use -
 PS Claim 20; Page 1198-1201; 1275pp; English.
 XX
 CC The present invention provides the protein and coding sequences of novel
 CC proteins from a variety of organisms, including human, dog, cat, horse,
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
 CC from the organism of interest. They can be used in diagnostics,
 CC forensics, gene mapping, identification of mutations, to assess
 CC biodiversity and for nutritional purposes. The present sequence is a
 CC protein of the invention.
 CC
 SQ Sequence 1299 AA;
 Query Match 91.7%; Score 6689; DB 22; Length 1299;
 Best Local Similarity 97.1%; Pred. No. 0;
 Matches 1257; Conservative 0; Mismatches 0; Indels 38; Gaps 1;
 QY 1 MAMKTPYLLLLLSVFIQOVSSODLSSCAGRCGEYSRPATNCPCYOHMECCPDF 60
 DB 1 mawktlpyllllsvfiqovssodlsscagrcgeysratncpcdynqhmeccpdf 60
 QY 61 KRVCTA-----AEVHNPTSPSSKKAP 82
 DB 61 krvtlaelscgrcfesfeirgedcdaqckkykccpdyesfaevhnpspskkap 120
 QY 83 PPSGASQTIKSTYKRSKPKPKKTKKYIESEETIEHVSSENOSSSSSSSSSTIWM 142
 DB 121 ppsgasqlkstkrspkpkkktkkyieeseeliehsvenqessssssssstliw 180
 QY 143 KIKSSKNSAANRELQKLVKNDKNNRTKKKPTKPPVVDAGSLDGDGVKVTTPDST 202
 DB 181 kikssksaanrelqklvndknnrtkkkppvvdagsldgdgvkvttpdst 240
 QY 203 TOHKKVSTSPKITTAKPINRPSLPNSDTSKETSLIVNKEITYETKETTNNKOTSDG 262
 DB 241 tqhkvstspkittakpinrpslppnsdtsketslrvnketvettltnkqtsldg 300
 QY 263 KEKTTSAKETOSIEKTSKADLAFTSKVLAKPTPKAETTKGPALUTTKKEPTTPPKEPAS 322
 DB 301 kekttsaketosiektsakdlaptskvlakptpkraetttkgpalttkkepttpkupas 360
 QY 323 TTPKEPTTITKSAVTPPKKEPAPTTTKSAPTTPKEPAPTTTKKEPAPTTTKKEPAPTTKEP 382
 DB 361 ttpkepttitsavtpkpkkepapttkesapttkkepapttkkepapttkkepapttkkep 420
 QY 383 APTTTSAPTTPKKEPAPTTPKKPAVTPPKKEPAPTTPKPEPTTPPKKEPAPTTPKK 442
 DB 421 aptttsaptpkkepapttkpkpavtpkpkkepapttkpepttpkkepapttkkepapttkp 480
 QY 443 EPAPTPAKKPAVTPPKKEPAPTTTKESPTTPPKKEPAPTTTSAPTTPKEPAPT 502
 DB 481 epaptpakpavtpkpkkepapttkespttpkkepapttkesapttkkepapttk 540
 QY 503 TTKSAPTPKKEPSPTTPKEPAPTTPKKPAVTPPKKAPATTPPKKEPAPTTPKK 562
 DB 541 tksaptpkkespttpkepapttkpkpavtpkpkapattppkkepapttkkepapttkk 600
 QY 563 APTAPKKEPAPTTPKETAPTTPKKLTPTPPEKLAPTTPEKPAVTPPEELAPTTPEEPPTT 622

OY	1223	EKKRGSGIDOOIYYIKOEPPVOKCGERRRALANPVYGEMOVRRRRRERARIGSSOFTIRIOX	1282
Db	1261	fkkrgsgidooiyyikoeppvokcgerrralanhpyvgemqvrirrtterai9psqtlctltiqy	1320
OY	1283	SFAPRLAYQDKGVLHNEVKVSIILMRGLPNVTSAISLPIINRIKPDGYDYAFPSKDQYVNDV	1342
Db	1321	sparlayqdkgylhnevkvslilwrglpnvvtlsalsipnlrlkpdgydyafpskdqynldv	1380
OY	1343	PSTRTAATITRSQGTLISKWYNCP	1366
Db	1381	pstaraatltrsqtlskwynncp	1404
 RESULT 2			
ID	AAB60568	standard; Protein; 1404 AA.	
AC	AAB60568;		
DT	27-APR-2001	(first entry)	
DE	Human megakaryocyte stimulating factor (MSF, CACP).		
KX	Human; CACP protein; camptodactyly-arthropathy-coxa vara-pericarditis;		
KW	MSF; megakaryocyte stimulating factor; synovial lubricant;		
KM	chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;		
KW	antiarthritic.		
OS	Homo sapiens.		
PN	WO200107068-A1.		
PD	01-FEB--2001.		
PF	21-JUL--2000; 2000MO-US200002.		
PR	23-JUL-1999; 99US-0145328.		
PR	19-JUL-2000; 2000US-0145328.		
PA	(UYCA-) UNIV CASE WESTERN RESERVE.		
PI	Warman ML;		
DR	WPI, 2001-182721/18.		
PT	New composition comprising the camptodactyly-arthropathy-coxa		
PS	vara-pericarditis protein in combination with an anesthetic, useful for		
XX	treating osteoarthritis, or as lubricants of tissue and joints		
XX	Example 1: Page -: 34pp; English.		
CC	The invention relates to a method of treating osteoarthritis via the		
CC	administration of a composition comprising the camptodactyly-arthropathy-		
CC	coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.		
CC	The composition may further comprise a local anaesthetic. The composition		
CC	of the invention may be administered via intra-articular or intravenous		
CC	injection. The human CACP protein is identified in the invention as		
CC	being megakaryocyte stimulating factor (MSF). The gene encoding		
CC	CACP protein (MSF) is located on chromosome 1q25-31, and mutations in		
CC	this gene are responsible for the heritable disorder camptodactyly-		
CC	arthropathy-coxa vara-pericarditis, in which patients have synovial		
CC	hyperplasia without evidence of inflammation. CACP protein (MSF)		
CC	acts as a synovium lubricant, and can be used to lubricate tissue and		
CC	joints in the treatment of osteoarthritis. The composition may be		
CC	applied to reduce the symptoms of osteoarthritis (e.g., joint pain,		
CC	loss of range of movement or joint damage). The present sequence		
CC	represents human megakaryocyte stimulating factor (MSF, CACP protein).		
CC	Note: This sequence is not given in its entirety in figure 4 of the		
CC	specification, although a GenBank accession number was given. This		
CC	sequence was therefore obtained from Genbank (U70316).		
XX	Sequence 1404 AA:		

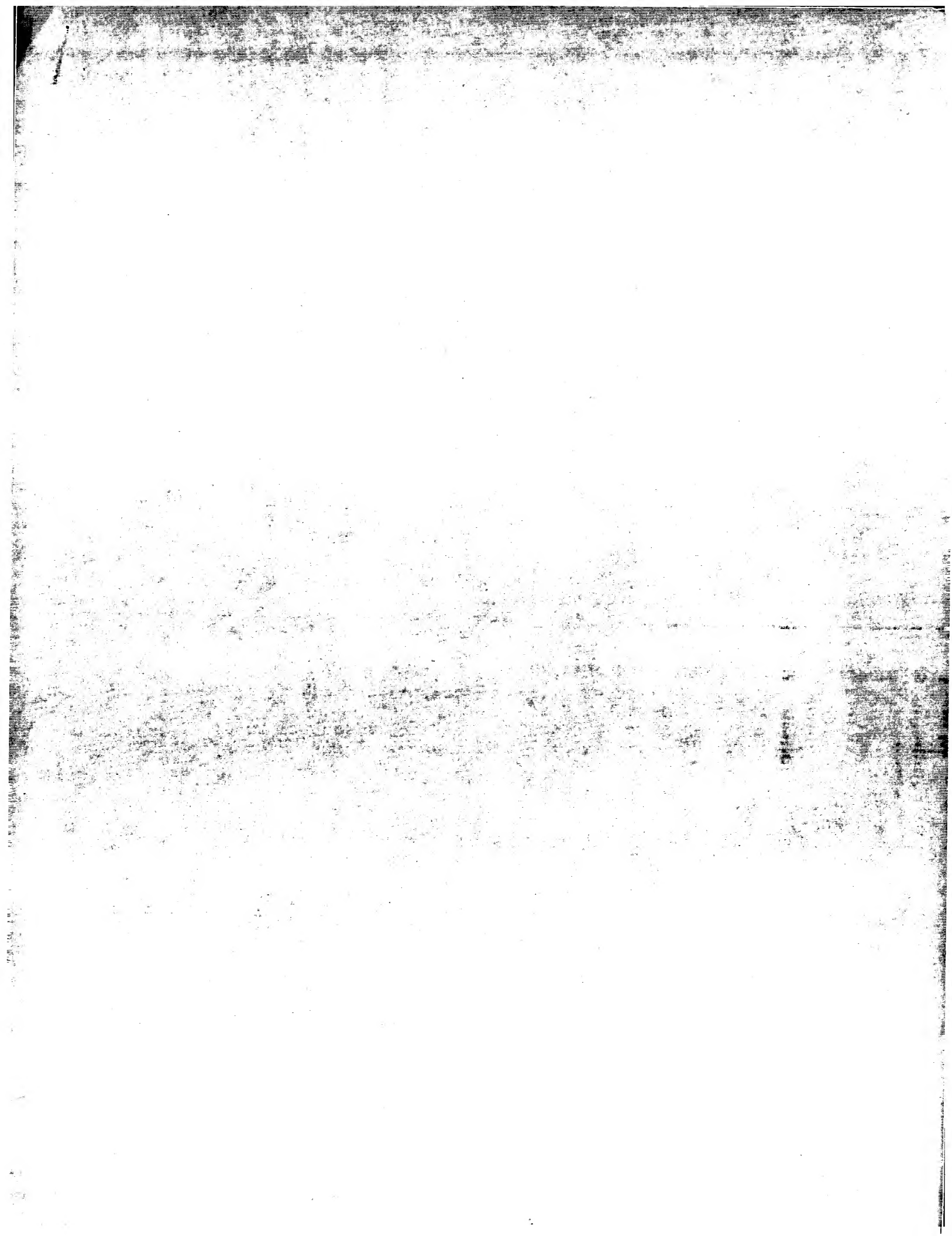
Query Match	99.6%	Score 7265	DB 22	Length 1404
Best Local Similarity	97.3%	Pred. Mismatches	0	
Matches 1366	Conservative	0	Indels 38	Gaps 1
QY	1	MAAKTPIVILLLLISFVFIQOOWSSODLSCAGCGGYSRDATCNDYNOAHMECCPDF	60	
DB	1	mawktlpllylllllsvfivqysagqlscagcggyardatcndcnoahymeccpdf	60	
QY	61	KRYCA-----AAVNPISPPSSKKAP	82	
DB	61	krycaelsackgrclesfargreccdaqckkydkccpdyesfcaevhnptspsskakap	120	
QY	83	PPSGASQITKSTTPKRSPPKPKKTKKVLSEETIEFHEVSEKOESSSSSSSSSTTAM	142	
DB	121	psgsagqtlkstltksrpkpkpkktkkyvleseetiehevsvengessssssssstlwm	180	
QY	143	KIKSSKNSAANRELQKLVKDNKNKRTKKPTPKPPVYDEAGSLDNGDFKVTPTDST	202	
DB	181	kikssksnaanelqkllkvkdnknrtkkppkppvvydeagsgldngdfkvtlptdst	240	
QY	203	TQHNNKYSIPKLTITAKPIPRESLPNSTPKETSLTVAKETTVKETTITTKKOSTG	262	
DB	241	tqnnkysipkltitakpimpripnsldstkseltvlnkettvetkeltlknqstlg	300	
QY	263	KKRTTSAKETSIIEKTSADLAPTSKVLAKPPKPAETTTKGPALTPKKEPTTPKEPAS	322	
DB	301	kertsaketstieksadlaptskvlnkppkpaetttkcpaltpkkepttpkepas	360	
QY	323	TPPKERTPTTISAPTTKEBPATTTKSAPTTPKEAPATTKEPATTTKEBPATTTKEP	382	
DB	361	tppkertpttisapttkebpatttksapttpkapattkebpatttkkebpatttkep	420	
QY	383	APTTKSAPTTPKEBPATTPPKKPAPPTPKBPATTPKEBPATTPPKBPATTPKEBPATTPK	442	
DB	421	apttksapttpkbpattppkppapptpkbpattpkbpattpkbpattpkbpattpk	480	
QY	443	EPAPLAPKKPAPTTKEBPATTPKBPATTTKEBSPTTKBPAPTTKSAPTTTKBPAPL	502	
DB	481	epaplpkppapttkbpattpkbpatttkespttkbpapttksaptttkbpaplk	540	
QY	503	TTKSAPTTPKEBSPTTKBPAPTTPKBPATTPKBPAPTTKBPAPTTKBPAPTTTKKP	562	
DB	541	tkksapttkbpatttkbpattpkbpattpkbpaptpkbpattpkbpattpkbp	600	
QY	563	APTAPKEPAPTTKEBPATTPPKKLPTTEBKLAPTTPKBPAPTTPEBPAPTT	622	
DB	601	aptapepapttketaptpkklpttpbeklapttpkbpapttpebpaptppt	660	
QY	623	PEBPATTTKKAAPNTPKBPAPTTPKBPAPTTPKBPAPTTPKETAATTTKGAATPLKBP	682	
DB	661	pebpatttkaaapntpkbpapttpkbpapttpkbpapttpketaattpkgatpplikbp	720	
QY	683	APTTKPKAPKELAPTTTKEPTSTSDKAPATTPKGTAPTTKEBPATTPKBPAPTTPG	742	
DB	721	aptpkpkapkelaptttkesptstsdkapattpkgatpttkebpattpkbpaptpbg	780	
QY	743	TAPTLKEBPAPTTPKBPAPKELAPTTTKGPTSTSSKAPATTPKETAATTPKBPAPTTK	802	
DB	781	taptlkebpattpkbpapakelaptttkgptstssdkapattpketaattpkebpattpk	840	
QY	803	KAPATTPERPPTTSVSPRTTTHKSEPTTLHKSDESTPELSAPPTPKALENSKEBGPV	862	
DB	841	kapattperppttsvsppttksepttlhksdetspelssapptpkalenskebgvpt	900	
QY	863	TATPAATPEEMTTAKDCTTERDLRTPTPTTAAPKMKETAATTEKTESKITATTTQV	922	
DB	901	tatpaatpeemttakdctterdlrtptpttaapkmketattekteskitattqy	960	
QY	923	NSTTODTTPKRTITLTKTTTLAPKVTTTKKTTTTEIMMKPEELAPKXORATNSKATTPK	982	
DB	961	tslttdttrpkrfttklttktlapkvtttkklttietimkpeelapkdoratnuskattpk	1020	

FT Region 1266..1331
 FT /label= Exon_X
 FT 1331..1373
 FT /label= Exon_XI
 FT 1373..1404
 FT /label= Exon_XII
 PN WO9213075-A.
 XX 06-AUG-1992.
 XX 17-JAN-1992; 92MO-US00433.
 XX
 XX 18-JAN-1991; 91US-0643502.
 XX 10-SEP-1991; 91US-0757022.
 XX (GEM) GENETICS INST INC.
 PA
 XX Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K;
 PI
 XX WPI: 1992-284660/34.
 DR N-PSDB; AAQ27223.
 XX
 PT New human mega-karyocyte stimulating factors - for treating
 PT immune deficiencies, cancer, exposure to radiation or drugs,
 PT bacterial and viral infections, etc.
 PS
 PS Claim 1, 2 and 3; Fig 1; 87pp; English.
 CC The sequence given is a full length translation from the megakaryocyte
 CC stimulating factor (MSF) precursor. The sequence covered by exons II,
 CC III and IV encodes megakaryocyte stimulating factor (MSF). This
 CC sequence is modified by the addition of an N-terminal sequence encoding
 CC a secretory leader, an initiating methionine proceeding exon II and a
 CC terminating codon following exon IV. The cDNA sequence given contains
 CC sequences derived from human megakaryocyte colony stimulating factor
 CC (meg-CSF). Exon I contains the initiating methionine, and encodes a
 CC classical mammalian protein secretion signal sequence. The sequence
 CC encoding the original meg-CSF includes exons II-IV and is thought to
 CC terminate in the region between amino acid residues 134 - 147. The
 CC primary transcript of this gene may be cleaved in different ways to
 CC yield a family of mRNA's each encoding a different MSF protein. Exons
 CC V and VI are thought to be related to the activity of the factor and
 CC are also implicated in the stability, folding and processing of the
 CC molecule. These exons are also thought to play a role in the observed
 CC synergy of MSF with other cytokines. Exons V - XII are believed to be
 CC implicated in the processing or folding of the appropriate structure of
 CC the resulting factor, ie. one or more of these exons may contain
 CC sequences which direct proteolytic cleavage, adhesion, organisation of
 CC the cellular matrix or extracellular matrix processing. Both naturally
 CC occurring and non-naturally occurring MSF's may be characterised by
 CC various combinations of alternatively spliced exons from this sequence,
 CC with the exons spliced together in differing orders to form different
 CC members of the MSF family.
 XX
 XX Sequence 1404 AA:

Query Match 99.6%; Score 7265; DB 13; Length 1404;
 Best Local Similarity 97.3%; Pred. No. 0;
 Matches 1366; Conservative 0; Mismatches 0; Indels 38; Gaps 1;

QY 1 MANKTLPYLLLVFVIOOVSSODLSCAGRCGEGYSRDATCNCDYNOCHMECCPDF 60
 DB 1 maktlpliyllllsvfviqvsqdlscagrcggyrsdatcncdncynqhmecpdf 60
 QY 61 KRVCCTA-----AEVHNPTSPSSKKAP 82
 DB 61 krcvtaelscjgrocfesfergedcdagckkydkccpdyesfcaehnpstpskkap 120
 QY 83 PRGASGTSITSTRSRSPKPNKKTKKVISEETEEHSHVSENOGSSSSSSSSSTIN 142
 DB 121 prsgaagtlsttkrtpkpnkktkkviesseteehshvseengessssssssstiw 180

QY 143 KIKSSKNSAANRELQKRLKVKDNKKRKKKPPKPPVVDAGSSGLDNGDFKVTPTDST 202
 DB 181 kiksksnsaanrelqkrlkvdnkkrkkkppkppvvdagsgjldngdfkvtptdst 240
 QY 203 TQHNKYSTSPKITTAKPIPNRSLPNSDTSKTSLSLVNKEFVETKTTTOKOSTDG 262
 DB 241 tqhnkystspkittakpnpnrslpnsdstsktslslvnketvetttkkqstldg 300
 QY 263 KEKTSAAKETOSIEKTSANDLAPTSKVLAKPPKATTTTGKALPTTPKEPTTPPEPAS 322
 DB 301 kektsaketosiektssandlaptskvlakppkatTTTGKALPTTPKEPTTPPEPAS 360
 QY 323 TTPKEPTPTTIKSAPTPPEKAPATTTKSAPTPPEKAPATTTKKAPATTTPKEP 382
 DB 361 ttpkeptpttiksaptppekapttkkapattTKSAPTPPEKAPATTTKKAPATTTPKEP 420
 QY 383 APPTTSAPTPPEKAPATTTPKAPATTTPKEAPATTTPKEPTTPPEKAPATTTPK 442
 DB 421 appttsaptppekapttttkapattTPKEAPATTTPKEPTTPPEKAPATTTPK 480
 QY 443 EPAP7APKRPAPTPPEKAPATTTPKEAPATTTPKEAPATTTPKAPATTTPK 502
 DB 481 epap7apkrpaptpppekapttttkapattTPKEAPATTTPKEAPATTTPKAPATTTPK 540
 QY 503 TTKSAPTPPEKSPPTTKKAPATTTPKEAPATTTPKAPATTTPKAPATTTPK 562
 DB 541 tksaptppeksppttkkapattTPKEAPATTTPKAPATTTPKAPATTTPK 600
 QY 563 AP7APKRPAPTPPEKAPATTTPKAPATTTPPEKAPATTTPPEKAPATTTPPEPT 622
 DB 601 ap7apkrpaptpppekapttttkapattTPPEKAPATTTPPEKAPATTTPPEPT 660
 QY 623 PEEPA7TPPKAAPNPKKAPATTTPKEAPATTTPKEAPATTTPKATTPPEKAPATT 682
 DB 661 peepa7tpkkaapnppkkapattTPKEAPATTTPKEAPATTTPKATTPPEKAPATT 720
 QY 683 APPTPKKAPKELAPTTTKEPTSTSDKAPATTTPKSTAPTTPEKAPATTTPK 742
 DB 721 aptpkkappkelaPTTTKEPTSTSDKAPATTTPKSTAPTTPEKAPATTTPK 780
 QY 743 TAPTTKEAPATPKKAPKELAPTTTKEPTSTSDKAPATTTPKAPATTTPKAPATT 802
 DB 781 tapttkkapatpkkappkelaPTTTKEPTSTSDKAPATTTPKAPATTTPKAPATT 840
 QY 803 KPA7TPPETPPTTSVSTPTTKEPTTIHKSDEESTPELSAPTPKALENSKEEGVPT 862
 DB 841 kpapttpeptpttsvstpttkepttiHKSDEESTPELSAPTPKALENSKEEGVPT 900
 QY 863 TKTPAATKPEMTTAAKDTTERDLRTTPETTAAPKMTKETAATTEKTSKITTATTOY 922
 DB 901 tktpaatkpemtttaakdttterdlrttpeTTAAPKMTKETAATTEKTSKITTATTOY 960
 QY 923 TSTPTODTTPKRTTTLTKTTTAPKVTTTKTTTITNMKPEETAPKXORATSKATTPK 982
 DB 961 tstptodtTPKRTTTLTKTTTAPKVTTTKTTTITNMKPEETAPKXORATSKATTPK 1020
 QY 983 POKPTAPKPKSTKPKKTPMRKPKTTPPRKMTSTMBELPPTSRIAEAMLOTTRPN 1042
 DB 1021 pokptapkpstkpkktpmrkpkTTPPRKMTSTMBELPPTSRIAEAMLOTTRPN 1080
 QY 1043 QTPNSKLVEVNPKSEDAGABGETPMMLLRPHVMEFVTPMDYDLRVPNOGIIINPMLS 1102
 DB 1081 qtpnsklvevnpksedagaagetpmlLRPHVMEFVTPMDYDLRVPNOGIIINPMLS 1140
 QY 1103 DETNINCGRPVDELITLRNGILVAFRGHFWMLSPSPSPSPARITTEWGISPDITVT 1162
 DB 1141 detnincgrpvdelitlrngilvafrgHFWMLSPSPSPSPARITTEWGISPDITVT 1200
 QY 1163 RCNCEGKTFEFPKDSOYWRFTNDIKDGYRKP1FKGGG1LGOIVAA1STAKYKMPESVY 1222
 DB 1201 rcncegkttfkdsgywrftndikdagypkpfkfgg1lgtivaalstakympevy 1260



aa6.ray

SEA ID NO: 1, 226
Ac NO: AAR26049,

728.782 Million cell updates/sec

Scoring table:

	BLOSUM62	Gap
Gap	10.0	Gapext 0.5

Scoring mode: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing:	Minimum Match	0%
	Maximum Match	100%

Listing first 45 summaries

```
Database : A_Geneseq_1101:*
1. /snp8/acadata/geneseq/geneseq/AA1980.DAT:*
```

1: /S1D58/gcgcatata/geneseq/genseq/AAL1980.DAT.*
2: /S1D58/gcgcatata/geneseq/genseq/AAL1980.DAT.*
3: /S1D58/gcgcatata/geneseq/genseq/AAL1982.DAT.*
4: /S1D58/gcgcatata/geneseq/genseq/AAL1983.DAT.*
5: /S1D58/gcgcatata/geneseq/genseq/AAL1984.DAT.*
6: /S1D58/gcgcatata/geneseq/genseq/AAL1985.DAT.*
7: /S1D58/gcgcatata/geneseq/genseq/AAL1986.DAT.*
8: /S1D58/gcgcatata/genseq/genseq/AAL1987.DAT.*
9: /S1D58/gcgcatata/genseq/genseq/AAL1988.DAT.*
10: /S1D58/gcgcatata/genseq/genseq/AAL1989.DAT.*
11: /S1D58/gcgcatata/genseq/genseq/AAL1990.DAT.*
12: /S1D58/gcgcatata/genseq/genseq/AAL1992.DAT.*
13: /S1D58/gcgcatata/genseq/genseq/AAL1993.DAT.*
14: /S1D58/gcgcatata/genseq/genseq/AAL1994.DAT.*
15: /S1D58/gcgcatata/genseq/genseq/AAL1995.DAT.*
16: /S1D58/gcgcatata/genseq/genseq/AAL1996.DAT.*
17: /S1D58/gcgcatata/genseq/genseq/AAL1997.DAT.*
18: /S1D58/gcgcatata/genseq/genseq/AAL1998.DAT.*
19: /S1D58/gcgcatata/genseq/genseq/AAL1999.DAT.*
20: /S1D58/gcgcatata/genseq/genseq/AAL2000.DAT.*
21: /S1D58/gcgcatata/genseq/genseq/AAL2001.DAT.*
22: /S1D58/gcgcatata/genseq/genseq/AAL2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	7265	99.6	1404	13	AA026049	MSF precursor. Sy
2	7265	99.6	1404	22	AA060568	Human megakaryocy
3	7265	99.6	1404	22	AA029773	Human megakaryocy
4	6684	91.7	1289	22	AA024322	Human EST encoded
5	3484	47.8	902	22	AA029778	Human MSF-derived
6	1936	26.5	452	16	AA080041	Human megakaryocy
7	1707.5	23.4	472	22	AA060569	Bovine MSF ortholo
8	1171	16.1	5179	22	AA024316	c899p predicted am
9	981	13.4	763	21	AA038942	Arabidopsis thalia
10	950	13.0	1664	21	AA043106	C. thermocellum OI
11	802	11.0	1325	22	AA003645	peptide Z2377 enco

12	112	115	9.8	7.3	18	AAH31852
11	640	612	8.8	4412	17	AAH53666
14	612	555.5	7.6	572	18	AAH31855
15	555.5	542	7.4	804	7	AAH60570
16	520.5	513	7.0	788	21	AAH54466
17	520.5	513	7.0	1837	21	AAH11726
18	506.5	506.5	6.9	744	9	AAH42875
19	506.5	506.5	6.9	1721	19	AAH82939
20	498.5	498.5	6.8	1721	21	AAH11727
21	498.5	497	6.8	182	12	AAH10872
22	496.5	496.5	6.8	2971	21	AAH41331
23	489.5	489.5	6.7	2972	22	AAH50363
24	489.5	489.5	6.7	3118	22	AAH50362
25	488.5	488.5	6.7	826	13	AAH26042
26	488.5	488.5	6.7	617	22	AAH61458
27	488	488	6.7	617	22	AAH04187
28	488	488	6.7	957	21	AAH59288
29	488	488	6.7	957	22	AAH424513
30	488	485	6.6	1127	22	AAH85541
31	476.5	476.5	6.5	652	9	AAH82974
32	467.5	467.5	6.4	511	22	AAH27312
33	467.5	467.5	6.4	511	22	AAH02607
34	467.5	467.5	6.4	511	22	AAH53408
35	467.5	467.5	6.3	2019	20	AAH517406
36	456	456	6.3	1012	20	AAH4160
37	454.5	454.5	6.2	378	12	AAH4162
38	450.5	450.5	6.2	1237	21	AAH91609
39	448	448	6.1	378	12	AAH4162
40	446.5	446.5	6.1	750	20	AAH05477
41	446.5	446.5	6.1	1532	21	AAH09454
42	444.5	444.5	6.1	3366	21	AAH82492
43	444.5	444.5	6.1	2665	22	AAH44533
44	443.5	443.5	6.1	2665	22	AAH62950

ALIGNMENTS

RESULT 1

ID AAR26049 standard; Protein; 1404 AA-

AC AAR26049;

DT 02-FEB-1993 (first entry)

DE MSF precursor.

XX Megakaryocyte colony stimulating factor; secretion signal; meg-CSF
KM stability; proteolytic cleavage; adhesion; alternative splicing.
KM

OS Synthetic.

FH	Key	Location/Qualifiers
1	36	

FT	/label= Exon_I
FT	

FT	label= Exon_1
107	

ET	label= Exon_1
107	157

Exon	Label	Exon	Label
1	Exon_1	1	Exon_1
2	Exon_2	2	Exon_2
3	Exon_3	3	Exon_3
4	Exon_4	4	Exon_4
5	Exon_5	5	Exon_5
6	Exon_6	6	Exon_6
7	Exon_7	7	Exon_7
8	Exon_8	8	Exon_8
9	Exon_9	9	Exon_9
10	Exon_10	10	Exon_10
11	Exon_11	11	Exon_11
12	Exon_12	12	Exon_12
13	Exon_13	13	Exon_13
14	Exon_14	14	Exon_14
15	Exon_15	15	Exon_15
16	Exon_16	16	Exon_16
17	Exon_17	17	Exon_17
18	Exon_18	18	Exon_18
19	Exon_19	19	Exon_19
20	Exon_20	20	Exon_20
21	Exon_21	21	Exon_21
22	Exon_22	22	Exon_22
23	Exon_23	23	Exon_23
24	Exon_24	24	Exon_24
25	Exon_25	25	Exon_25
26	Exon_26	26	Exon_26
27	Exon_27	27	Exon_27
28	Exon_28	28	Exon_28
29	Exon_29	29	Exon_29
30	Exon_30	30	Exon_30
31	Exon_31	31	Exon_31
32	Exon_32	32	Exon_32
33	Exon_33	33	Exon_33
34	Exon_34	34	Exon_34
35	Exon_35	35	Exon_35
36	Exon_36	36	Exon_36
37	Exon_37	37	Exon_37
38	Exon_38	38	Exon_38
39	Exon_39	39	Exon_39
40	Exon_40	40	Exon_40
41	Exon_41	41	Exon_41
42	Exon_42	42	Exon_42
43	Exon_43	43	Exon_43
44	Exon_44	44	Exon_44
45	Exon_45	45	Exon_45
46	Exon_46	46	Exon_46
47	Exon_47	47	Exon_47
48	Exon_48	48	Exon_48
49	Exon_49	49	Exon_49
50	Exon_50	50	Exon_50
51	Exon_51	51	Exon_51
52	Exon_52	52	Exon_52
53	Exon_53	53	Exon_53
54	Exon_54	54	Exon_54
55	Exon_55	55	Exon_55
56	Exon_56	56	Exon_56
57	Exon_57	57	Exon_57
58	Exon_58	58	Exon_58
59	Exon_59	59	Exon_59
60	Exon_60	60	Exon_60
61	Exon_61	61	Exon_61
62	Exon_62	62	Exon_62
63	Exon_63	63	Exon_63
64	Exon_64	64	Exon_64
65	Exon_65	65	Exon_65
66	Exon_66	66	Exon_66
67	Exon_67	67	Exon_67
68	Exon_68	68	Exon_68
69	Exon_69	69	Exon_69
70	Exon_70	70	Exon_70
71	Exon_71	71	Exon_71
72	Exon_72	72	Exon_72
73	Exon_73	73	Exon_73
74	Exon_74	74	Exon_74
75	Exon_75	75	Exon_75
76	Exon_76	76	Exon_76
77	Exon_77	77	Exon_77
78	Exon_78	78	Exon_78
79	Exon_79	79	Exon_79
80	Exon_80	80	Exon_80
81	Exon_81	81	Exon_81
82	Exon_82	82	Exon_82
83	Exon_83	83	Exon_83
84	Exon_84	84	Exon_84
85	Exon_85	85	Exon_85
86	Exon_86	86	Exon_86
87	Exon_87	87	Exon_87
88	Exon_88	88	Exon_88
89	Exon_89	89	Exon_89
90	Exon_90	90	Exon_90
91	Exon_91	91	

```

/label= Exon_V
200 1141

```

```

      /label= Exon_V
      1411..1166
      Region
      ET
      ET

```

Region	ET	ET
1166..1212	/label= each_v	

Region	FT	FT
1213..1266	1213..1266	1213..1266

F

	RESULT	1
ID	AAR26049	
AC	AAR26049	standard; Protein; 1404 AA
XX	02-FEB-1993	(first entry)
DY		
XX		
DE	MSF precursor.	
XX		
KW	Negatively charged colony stimulating factor stability; proteolytic cleavage; ad	
KW		
OS	Synthetic.	
XX		
FH	Key	Location/Qualifiers
FT	Region	1..26
FT		/label= Exon_I
FT	Region	26..67
FT		/label= Exon_II
FT	Region	67..107
FT		/label= Exon_III
FT	Region	107..157
FT		/label= Exon_IV
FT	Region	157..200
FT		/label= Exon_V
FT	Region	200..1141
FT		/label= Exon_VI
FT	Region	1411..1166
FT		/label= Exon_VII
FT	Region	1166..1212
FT		/label= Exon_VIII
FT	Region	1212..1266
FT		/label= Exon_IX

Oy	1086	FFGGGSGTQOYITKCEPVOKGCRBPALNVPYEGMTQVRRRRFEAISPSCHTIRIQY	1145
Db	1261	ffkrgsgdqylykqepqkcpgrirpalnhyrgentqyrrrrferaiqpsqhtlrlqy	1320
Oy	1146	SPARLAYODKGLHNEKRVYSILMRGLPNVVTSAISLSPNPKRDKGYDYAFSKDQYINIDV	1205
Db	1321	sparlqydkqylhnevksllwrlplmwrttsaislplnlrkpdyqyafskdqynldv	1380
Oy	1206	PSRTAPAAITRRSGOTLSKRWYNCP	1229
Db	1381	psrtaralttrsgqlskrwynncp	1404
RESULT	2		
AAB00568			
ID	AAB0568	standard; Protein; 1404 AA.	
AC	AAB0568;		
DT	27-APR-2001	(first entry)	
DE	Human megakaryocyte stimulating factor (MSF, CACP).		
KW	Human; CACP protein; campitodactylly-arthropathy-coxa vara-pericarditis;		
KW	MSF; megakaryocyte stimulating factor; synovial lubricant;		
KW	chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;		
KW	antiarthritic.		
OS	Homo sapiens.		
PN	WO200107068-A1.		
PD	01-FEB-2001.		
PE	21-JUL-2000; 20RQWO-US20002.		
PR	23-JUL-1999; 99US-0145328.		
PR	19-JUL-2000; 2000US-0145328.		
PA	(UT2A-) UNIT CASE WESTERN RESERVE.		
PI	Wayman ML;		
DR	WP1; 2001-182721/18.		
PT	New composition comprising the campitodactylly-arthropathy-coxa		
PT	vara-pericarditis protein in combination with an anesthetic, useful for		
PT	treating osteoarthritis, or as lubricants of tissue and joints		
PS	Example 1; Page 7; 34pp; English.		
XX	The invention relates to a method of treating osteoarthritis via the		
XX	administration of a composition comprising the campitodactylly-arthropathy-		
XX	coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.		
XX	The composition may further comprise a local anesthetic. The composition		
XX	of the invention may be administered via intra-articular or intravenous		
XX	injection. The human CACP protein is identified in the invention as		
XX	being megakaryocyte stimulating factor (MSF). The gene encoding		
XX	CACP protein (MSF) is located on chromosome 1q25-31, and mutations in		
XX	this gene are responsible for the heritable disorder campitodactylly-		
XX	arthropathy-coxa vara-pericarditis, in which patients have synovial		
XX	hyperplasia without evidence of inflammation. CACP protein (MSF)		
XX	acts as a synovium lubricant, and can be used to lubricate tissue and		
XX	joints in the treatment of osteoarthritis. The composition may be		
XX	applied to reduce the symptoms of osteoarthritis (e.g., joint pain,		
XX	loss of range of movement or joint damage). The present sequence		
XX	represents human megakaryocyte stimulating factor (MSF, CACP protein).		
XX	Note: This sequence is not given in its entirety in figure 4 of the		
XX	specification, although a Genbank accession number was given. This		
XX	sequence was therefore obtained from Genbank (U00316).		
XX	Sequence	1404 AA;	

[illegible]

Human SRCP. Homo
P. yoeili SSP2 ant
Peptide #2892 enco
Peptide #2869 enco
Human MUC1 polype
C9orf predicted am
Human protein sequ
Bioscience precu
Peptide #317 enco
Peptide #1349 enco
Peptide #1289 enco
Human 07CG7 gene
PRP 378. Titlun
Peptide #957 enco
Peptide #967 enco
Peptide #941 enco
Human ORF ORF225
PRP encoded by cl
C. albicans Rb11-1
Human atrophin-1


```

FT FT Region 1266..1331 /label=- Exon_X
FT FT Region 1331..1373
FT FT Region /label=- Exon_XI
FT FT Region 1373..1404
XX XX /label= Exon_XII
PN M09213075-A.
PD 05-AUG-1992.
PF 17-JAN-1992; 92MO-USO0433.
XX 18-JAN-1991; 91US-0643502.
PR 10-SEP-1991; 91US-0757022.
XX (GENY ) GENETICS INST INC.
PI Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K;
DR WPI; 1992-284660/34.
DR N-PSDB; AAO27223.
XX New human mega-karyocyte stimulating factors - for treating
PT immune deficiencies, cancer, exposure to radiation or drugs,
PT bacterial and viral infections, etc.
XX Claim 1, 2 and 3; Fig 1; 87Pp; English.
XX The sequence given is a full length translation from the megakaryocyte
CC stimulating factor (MSF) precursor. The sequence covered by exons II,
CC III and IV encodes megakaryocyte stimulating factor (MSF). This
CC sequence is modified by the addition of an N-terminal sequence encoding
CC a secretory leader, an initiating methionine preceding exon II and a
CC terminating codon following exon IV. The cDNA sequence given contains
CC sequences derived from human megakaryocyte colony stimulating factor
CC (meg-CSF). Exon I contains the initiating methionine, and encodes a
CC classical mammalian protein secretion signal sequence. The sequence
CC encoding the original meg-CSF includes exons II-IV and is thought to
CC terminate in the region between amino acid residues 134 - 147. The
CC primary transcript of this gene may be cleaved in different ways to
CC yield a family of mRNAs each encoding a different MSF protein. Exons
CC V and VI are thought to be related to the activity of the factor and
CC are also implicated in the stability, folding and processing of the
CC molecule. These exons are also thought to play a role in the observed
CC synergy of MSF with other cytokines. Exons V - XII are believed to be
CC implicated in the processing or folding of the appropriate structure of
CC the resulting factor, ie. one or more of these exons may contain
CC sequences which direct proteolytic cleavage, adhesion, organisation of
CC the cellular matrix or extracellular matrix processing. Both naturally
CC occurring and non-naturally occurring MSFs may be characterised by
CC various combinations of alternatively spliced exons from this sequence.
CC With the exons spliced together in differing orders to form different
CC members of the MSF family.
SQ Sequence 1404 AA:
Query Match 99.6%; Score 7291.5; DB 13; Length 1404;
Best Local Similarity 96.9%; Pred. No. 0;
Matches 1361; Conservative 0; Mis-matches 0; Indels 43; Gaps 1
DY 1AAMKTLPIYLILLISVFVIOGVSDLDLSACARCGEYSRDATNCNDYACOHMECCPDF 60
Db 1MAWKILPIYIIILSVFYLVGVSQGLSCAGRCGSGYRDTCTCNDYCNHYMECCPAF 60
QY KRVCFAELSCNCRCESEFERECODADCKKYDKCCPDYEFSCAEVNHPSPSSKKAP 120
Db 61 KRVCAELSCNCRCESEFERECODADCKKYDKCCPDYEFSCAEVNHPSPSSKKAP 120
QY 121 PPGSASQTIKSTTKSPKPNNKKTTRKVIESSEITE----- 156
Db 121 PPGSASQTIKSTTKSPKPNNKKTTRKVIESSEITE----- 156

```

QY	157	-----VKDNKNKNTKRRKPPKPPVVDASGNDGNCFKTTGPDTS197
Db	181	k4ksksnseanrelqk4k1kxdknhrtkkcpkppvvdasgslangdfkctcdtc240
QY	198	TOHNKVSITSPIKATTAKDINRPSLIPNSDSKSTSLVKNKEIVVEKETTNTTKQSTDG257
Db	241	tqhnkvsstprlttakp1nprpslpposdstsketsltnvkettvettkettlnkqstgd300
QY	258	KEKTTSAKENOSIKNTSANKDLAPTSKYLAPTKAEITTKGPMLTTPKEPTPTTPKEPNS317
Db	301	kektsakeitsetslektakolaptskylaptkaeittkgpal1tkpkepttkepas360
QY	318	TTPEEPPTTTKSAPTTPKBPAPTTTSSATTPKEBPATTTKBPATTPPEBPATTTTKBP377
Db	361	ttpeeppttkksapttkbpaptttssattpkebpatttkbpattppkebpatttkbp420
QY	378	APTTSKAPTTPKBPAPTTTRKBPATTPTPKBPAPTTPEPTTPTPKBPAPTTKBPATTPRK437
Db	421	apttksapttkbpaptttrkbpapttptpkbpapttpepttptpkbpapttkbpapttkrk480
QY	438	EPAPLAKKBPAPTTTPKBPAPTTTPKBPAPTTTKBSPPTTPKEBPAPTTTKSAPTTTKBPAP497
Db	481	epapltakpkbpaptttkbpaptttkbpaptttkbepspttkbpapttksaptttkbpap540
QY	498	TTKSAPTTPKBPSTTTTKBPAPTTTPKBPAPTTTKBPAPTTTPKBPAPTTTPPEBPATTTTKKP557
Db	541	tkksapttkbpespsttkkbpaptttkbpaptttkbpaptttkbpaptttkbpaptttkbp600
QY	558	APTAPKEBPAPTTPEIAPTTTPPKLTPTPPKLAPTTPEKAPAPTTPEIAPTTPEEPPTPT617
Db	601	aptapkebpapttpeiaptttppkltptppkltapttpekapapttpeiapttpeepptpt660
QY	618	PEBPAPTTPKAAPNTPKBPAPTTTPKBPAPTTTPKBPAPTTPKETAPTTPKGAAPTTLAKP677
Db	661	peebpapttpkaapntpkbpaptttkbpaptttkbpaptttkbpaptttkbpaptttkbp720
QY	678	APTTPKRPAPELAPTTTKBPTSTTSKAPAPTTPKGAAPTTPKBPAPTTTPKBPAPTTTPKG737
Db	721	aptpkpkrapelaptttkbptsttskapaapttpkgaptttkbpaptttkbpaptttkbp780
QY	738	TAPTTLKEBPAPTTTPKKEBPAPELAPTTTKGPTSTTSKAPAPTTPKEBPAPTTTPKBPAPTPK797
Db	781	tapttlkebpaptttpkkepapelaptttkgptsttsdkapaptttkbpaptttkbpaptttk840
QY	798	KPAPTTPPEPTTSSEVSTPTTKKEPTTIHKSDESTPELSABTPKALENSEKEBGVPT857
Db	841	kpaapttpepttsevspttkkepttihsdestpelsabtpkalenskebgvpt900
QY	858	TKTPRAATPEMTTAKOKTTEROULTPTBETTTAAPKTKTATTTEKTBESKTATTATTOY917
Db	901	tktpaatpemttaoktcterdlrttpeettaapkmktelatttektesktattattoy960
QY	918	TSTTPQDTPPKITTLTKTTLAPVTVTTKTITTTTELHNKPEETAPRKDQATNSKATTPK977
Db	961	tstttqdtppkittlktlclttclapkvttcttkctittelaimpbeetaprkatactskatr902
QY	978	POKPTKAPKKEPTSTKKEPTMRVRKKEPTTPPKMTSTMBELNPTSIKIAEAMQTTTRPN1033
Db	1021	pokptkapkktstkkkptkmpvrvkrcpttpkmtsimpelnpstareleamqtttrpn108
QY	1038	QTPNSKLVENVKESDAGGABGETPHMLNRPVMEVPTPDMDYLPKRVNOGIINPMUS1099
Db	1081	qtprnsklvenvpkedaggaegecprhmlrphvfmpevtpdmdy1pvrpqg11nprms114
QY	1098	DETNNCKGKPRDGLTTLRNGNLVAFRGHYTWMLSPPSPBPARKRIEWMG1PSPIDTWT115
Db	1141	detnncgkprdglttlrngnlvafrghytwm1spspbparrtievwg1psp1dtvt120
QY	1158	RCONEGKTFEFKDSQVWRFTNDIKIDACQPKPIFKGFGGLGOLIVALSATKYNMPEASY121
Db	1201	rcnegrkftfkdsqwrftndikidagypkpiikfgfvg1tqglvaalsatkynmpeasy126

QY 1218 FFRGGSIQOYIYKQEPVOKCPGRPALNYPVYGEOTVRRRFRERAIQSPQTHIRIQY 1277
 DB 1261 ffrggsiaglykykgepvokcpgrpalnypvgemtyqrrrrrferaispsqthirirly 1320
 QY 1278 SPARLAQOKGVILHNEVKVSIIMRGCLPNVNTAISLPIIRKDGIDYAFSSQDYNNIDV 1337
 DB 1321 sparlayqdkgvilhevkvsilwrglpinvtalsipnirkpdydyafsfkxdqynidv 1380
 QY 1338 PSTARAITRRSGQTLRSKVMNCP 1361
 DB 1381 pstaraitrrsgqtlrskvwnpc 1404

RESULT 2

AA0568
 ID AAB60568 standard; Protein: 1404 AA.

AC AAB60568;

DT 27-APR-2001 (first entry)

DE Human megakaryocyte stimulating factor (MSF, CACP).

KW Human: CACP protein; campodactylly-arthropathy-coxa vara-pericarditis;
 MSF; megakaryocyte stimulating factor; synovial lubricant;

KM chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;
 arthritis; etc.

KX Homo sapiens.

OS Homo sapiens.

PN W0200107068-AA.

PD 01-FEB-2001.

PF 21-JUL-2000; 2000NO-US20002.

PR 23-JUL-1999; 98US-0145328.

PR 19-JUL-2000; 2000US-0145328.

XX (DYCA-) UNIV CASE WESTERN RESERVE.

PA Weiman ML;

PI WPI: 2001-182721/18.

PT New composition comprising the campodactylly-arthropathy-coxa
 vara-pericarditis protein in combination with an anesthetic, useful for
 treating osteoarthritis, or as lubricants of tissue and joints

PT Example 1: Page -: 34pp; English.

XX The invention relates to a method of treating osteoarthritis via the
 administration of a composition comprising the campodactylly-arthropathy-
 coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.

CC The composition may further comprise a local anesthetic. The composition
 of the invention may be administered via intra-articular or intravenous
 injection. The human CACP protein is identified in the invention as
 being megakaryocyte stimulating factor (MSF). The gene encoding
 CACP protein (MSF) is located on chromosome 1q25-31, and mutations in
 this gene are responsible for the heritable disorder campodactylly-
 arthropathy-coxa vara-pericarditis, in which patients have synovial
 hyperplasia without evidence of inflammation. CACP protein (MSF)
 acts as a synovium lubricant, and can be used to lubricate tissue and
 joints in the treatment of osteoarthritis. The composition may be
 applied to reduce the symptoms of osteoarthritis (e.g., joint pain,
 loss of range of movement or joint damage). The present sequence
 represents human megakaryocyte stimulating factor (MSF, CACP protein).

CC Note: This sequence is not given in its entirety in figure 4 of the
 specification, although a Genbank accession number was given. This
 sequence was therefore obtained from Genbank (U70316).

CC Sequence was therefore obtained from Genbank (U70316).

CC Sequence was therefore obtained from Genbank (U70316).

CC Sequence was therefore obtained from Genbank (U70316).

CC Sequence was therefore obtained from Genbank (U70316).

CC Sequence was therefore obtained from Genbank (U70316).

CC Sequence was therefore obtained from Genbank (U70316).

CC Sequence was therefore obtained from Genbank (U70316).

CC Sequence was therefore obtained from Genbank (U70316).

CC Sequence was therefore obtained from Genbank (U70316).

CC Sequence was therefore obtained from Genbank (U70316).

Query Match 99.6%; Score 7291.5; DB 22; Length 1404;
 Best Local Similarity 96.9%; Pred. No. 0;
 Matches 1361; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY 1 MAMTLPIYLLLLSVFIIQOVSSODLSGCGRGEGSRATONCYNCHYMCCGDF 60
 DB 1 MAMTLPIYLLLLSVFIIQOVSSODLSGCGRGEGSRATONCYNCHYMCCGDF 60
 QY 61 KVCYTAELSCGRCFESFEREGECDAOCCKYDKCCPDYEFSCAEYHNPSPSSKAP 120
 DB 61 KVCYTAELSCGRCFESFEREGECDAOCCKYDKCCPDYEFSCAEYHNPSPSSKAP 120
 QY 121 PPSGASQTIKSTKRSKRPKKTKKYVISEELTE----- 156
 DB 121 PPSGASQTIKSTKRSKRPKKTKKYVISEELTE----- 156
 QY 157 -----VADNKNRTRKKKPKPKPVYDEAGSLGNGDFKVTPTST 197
 DB 181 KIKESKNSAANRELQKLLKDKNNKTKKPKPVYDEAGSLGNGDFKVTPTST 240
 QY 198 TQHNKVSSTPKITTAAPINPNSLPNSDTSKETSILVNNKETTVEKETTNNKQTSIDG 257
 DB 241 TQHNKVSSTPKITTAAPINPNSLPNSDTSKETSILVNNKETTVEKETTNNKQTSIDG 300
 QY 258 KKTTSKKNOSKETSASAKDLAPSKYLAAPTPKAETTTGPAITTRKEPTTPPKEPAS 317
 DB 301 KKTTSKKNOSKETSASAKDLAPSKYLAAPTPKAETTTGPAITTRKEPTTPPKEPAS 360
 QY 318 TTPKEPTTPPKEPASPTTPPKEPAPPTTTPPKEPAPPTTTPPKEPAPPTTTPPK 377
 DB 361 TTPKEPTTPPKEPASPTTPPKEPAPPTTTPPKEPAPPTTTPPKEPAPPTTTPPK 420
 QY 378 APPTTKAPPTPKEPAPPTTPPKEPAPPTTTPPKEPAPPTTTPPKEPAPPTTTPK 437
 DB 421 APPTTKAPPTPKEPAPPTTPPKEPAPPTTTPPKEPAPPTTTPPKEPAPPTTTPK 480
 QY 438 EPAPTAKKRPAPPTPKEPAPPTTTPPKEPAPPTTTPPKEPAPPTTTPPKEPAPPT 497
 DB 481 EPAPTAKKRPAPPTPKEPAPPTTTPPKEPAPPTTTPPKEPAPPTTTPPKEPAPPT 540
 QY 498 TTKSAPPTPKEPAPPTTTPPKEPAPPTTTPPKEPAPPTTTPPKEPAPPTTTPK 557
 DB 541 TTKSAPPTPKEPAPPTTTPPKEPAPPTTTPPKEPAPPTTTPPKEPAPPTTTPK 600
 QY 558 APPTAKKRPAPPTPKEPAPPTTTPPKEPAPPTTTPPKEPAPPTTTPPKEPAPPT 617
 DB 601 APPTAKKRPAPPTPKEPAPPTTTPPKEPAPPTTTPPKEPAPPTTTPPKEPAPPT 660
 QY 618 PEEPAPPTPKAAPNPMPKEPAPPTPKEPAPPTTTPPKEPAPPTTTPPKEPAPPT 677
 DB 661 PEEPAPPTPKAAPNPMPKEPAPPTPKEPAPPTTTPPKEPAPPTTTPPKEPAPPT 720
 QY 678 APPTPKKAPKELAPPTTKETSTSDKAPPTPKGAPPTTPPKEPAPPTPKG 737
 DB 721 APPTPKKAPKELAPPTTKETSTSDKAPPTPKGAPPTTPPKEPAPPTPKG 780
 QY 738 TAPPTLKEPAPPTPKKAPKELAPPTTKGPTSTSDKAPPTPKGAPPTTPPK 797
 DB 781 TAPPTLKEPAPPTPKKAPKELAPPTTKGPTSTSDKAPPTPKGAPPTTPPK 840
 QY 798 KPAPTTPPTPPTSEVSTTPPKETTHKSPDESPELSAPPTPKALENSKEGVP 857
 DB 841 KPAPTTPPTPPTSEVSTTPPKETTHKSPDESPELSAPPTPKALENSKEGVP 900
 QY 858 TKPPAATKPPMTTAAOKTTERDLRTTPETTTAAPKREKATTTKTESKTTATTTQV 917
 DB 901 TKPPAATKPPMTTAAOKTTERDLRTTPETTTAAPKREKATTTKTESKTTATTTQV 960
 QY 918 TSTTODTPEFTTTLTKTTTLAPKVTYTTTKTTTITIMKPEETAKPKDRAVNSKATTPK 977
 DB 961 TSTTODTPEFTTTLTKTTTLAPKVTYTTTKTTTITIMKPEETAKPKDRAVNSKATTPK 1020

EEA ID 1, AAR
Ac No: AAR-26049, A-Genesys-1101

Myobacterium tube
Sequence 91/101742
Myobacterium tube
Sequence of the Pa
Aminl acid sequenc
Protein encoded by
Amino acid sequenc
Cryosporidium pa
Bioactive precur
MSF-K130 Synthet
Human ORF9 ORF95
Human SRCAP. Homo
Human SRCAP. Homo
Portion of Crypto
P. yoelii SS2 ant
Peptide #2892 enco
Peptide #2863 enco
Human MUC11 polype
C900P predicted am
Cryosporidium pa
Human protein sequ
Bioactive precur
Peptide #1317 enco
Peptide #1319 enco
Peptide #1289 enco
C. albicans Rbt1 P
PRP 378. Tricolum
PRP encoded by cld
Caenorhabditis elat
Caenorhabditis elat
Japanese sea muss
Human 07C37 gene
HYFP-DEVD-NAP4-EBB
Bifunctional caspae

FT Region 1266..1331
FT /label= Exon_X
FT 1331..1373
FT /label= Exon_XI
FT 1373..1404
FT /label= Exon_XII
XX WQ9213075-A.
XX
XX 06-AUG-1992.
XX
XX 17-JAN-1992; 92WO-US00433.
XX
XX 18-JAN-1991; 91US-0643502.
XX 10-SEP-1991; 91US-0757022.
XX
XX (GENY) GENETICS INST INC.
XX
XX Clark SC, Gesner TG, Hewlck RM, Jacobs K, Turner K;
XX WPI: 1992-284660/34.
XX N:PSDB:AAQ27223.
XX
XX New human mega-karyocyte stimulating factors - for treating
XX immune deficiencies, cancer, exposure to radiation or drugs,
XX bacterial and viral infections, etc.
XX
XX Claim 1, 2 and 3; Fig 1; 87pp: English.
XX
XX The sequence given is a full length translation from the megakaryocyte
XX stimulating factor (MSF) precursor. The sequence covered by exons II,
XX III and IV encodes megakaryocyte stimulating factor (MSF). This
XX sequence is modified by the addition of an N-terminal sequence encoding
XX a secretory leader, an initiating methionine preceding exon II and a
XX terminating codon following exon IV. The cDNA sequence given contains
XX sequences derived from human megakaryocyte colony stimulating factor
XX (meg-CSF). Exon I contains the initiating methionine, and encodes a
XX classical mammalian protein secretion signal sequence. The sequence
XX encoding the original meg-CSF includes exons II-IV and is thought to
XX terminate in the region between amino acid residues 134 - 147. The
XX primary transcript of this gene may be cleaved in different ways to
XX yield a family of mRNA's each encoding a different MSF protein. Exons
XX V and VI are thought to be related to the activity of the factor and
XX are also implicated in the stability, folding and processing of the
XX molecule. These exons are also thought to play a role in the observed
XX synergy of MSF with other cytokines. Exons V - XII are believed to be
XX implicated in the processing or folding of the appropriate structure of
XX the resulting factor, ie. one or more of these exons may contain
XX sequences which direct proteolytic cleavage, adhesion, organisation of
XX the cellular matrix or extracellular matrix processing. Both naturally
XX occurring and non-naturally occurring MSF's may be characterised by
XX various combinations of alternatively spliced exons from this sequence,
XX with the exons spliced together in differing orders to form different
XX members of the MSF family.
XX
XX Sequence 1404 AA:
XX
Query Match 99.2%; Score 7007.5; DB 13; Length 1404;
Best Local Similarity 93.4%; Pred. No. 0;
Matches 131; Conservative 0; Mismatches 0; Indels 93; Gaps 1;
QY 1 MAMTTLPTLLLSVYIQVSSQDSSCAGRCGEGSRDATCNCDYNCHYMCCPDF 60
DB 1 mawtllptllllsvfvlqvgssqdlsscagrcgsgysratcdncdyncqhmccpdf 60
QY 1 KRYVTAESCGRGFESEEREGDCDAQCKKXDKCCPDYSSFAE----- 106
DB 1 krvytaescgrgfefergedcdagckkxdkccpdyesfaevhnptspsskkap 120
QY 107 ----- 106
DB 121 ppsgasqlktcttrspkpkpkkttkkylaseelteehtsvsengessssssssstliw 180

QY 107 -----VKDNKKNRKKRRKPPVNDAGSGLDNGDFKVTTPNST 147
DB 181 kksksksaanrelqkklxvdknkrkxkpkpvpvdaagslndgdkkvtpdst 240
QY 148 TOHNKVSSTPKITTAKPINRPISLPNSDTSKEPISLVNKEVTEFKETTTNKTSDG 207
DB 241 tqhkvstspkltaklpinrpislpnsdtskelslvnketvettcttnqtsdsg 300
QY 208 KEKTSKAKETQSIEKTSKADLAPTSKVLAKPRAETTTKGALVTPKEPPTPKEPAS 267
DB 301 kektsaketqsiektsakdlaptskvlakpbraetttkgalvtpkepptpkepas 360
QY 268 TTPKEPPTTIKSAPTTPKBPAPTTTSAPTTPKBPAPTTTKEPAPTTTKEP 327
DB 361 ttpkeppttiksaptpkbpapttttsaptpkbpaptttkepaptttkepaptttke 420
QY 328 APPTTKSAPPTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 387
DB 421 appttksapptpkbpaptpkbpaptpkbpaptpkbpaptpkbpaptpkbpaptpk 480
QY 388 EPAPTAAPKAPATTPKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP 447
DB 481 epaptaapkapattpkbpapttkbpapttkbpapttkbpapttkbpapttkbpap 540
QY 448 TTKSAPTTKBPSTTTKEBPATTPKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 507
DB 541 tksapttkbpstttkebpattpkbpapttkbpapttkbpapttkbpapttkbpap 600
QY 508 APATPKBPAPTTPKBPATTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBP 567
DB 601 apatpkbpaptpkbpaptpkbpaptpkbpaptpkbpaptpkbpaptpkbpaptpk 660
QY 568 PEEPAPTPPAAANPNTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP 627
DB 661 peepaptppaaanpntkbpapttkbpapttkbpapttkbpapttkbpapttkbpap 720
QY 628 APPTPKPAKELAPTTTKEPSTSTSDKPAPTTPKGAPTTPKBPAPTTKBPAPTTK 687
DB 721 aptpkpapelaptttkepststsdkpapttpkgapttpkbpapttkbpapttkbp 780
QY 688 TAPPTLEBPATTPKBPAPTELAPTTKGPSTSDKPAPTTPKELAPTTKBPAPTTK 747
DB 781 tapptlebpattpkbpaptelapttkgpstsdkpapttpkelapttkbpapttk 840
QY 748 KPAPTPETPPTTSEVSTPTTKEPPTIKSPDESPELSAEPKALENSPEEPVPT 807
DB 841 kpaptpetppttsevsppttkepptikspdespelasaepkalenspeepvpt 900
QY 808 TKTPPAATKPEMTTAKDKTERDLRTTPETTTAAPKMTKETATTEKTSKIRATTTQV 867
DB 901 tktpaatkpemttakdkterdlrttpeettaapkmketaattektskiratttqv 960
QY 868 TSTTTOOTTPFKITTLKTTTLAPKVYTTTKTITTTETIMNKPDEGTAKKDRATNSKATTP 927
DB 961 tstttoottpfkittllktttlapkvtttkitttetimnkpdegtakkdraztnakatpk 1020
QY 928 POKPTKAPKPTSTKRPKTPRVAKKPTTPPKKSTMPBELNSTIAEAMLOTTTRPN 987
DB 1021 pokptkapkptstkrpktprvakkpttpkkstmpbelnstiaeamlootttrpn 1080
QY 988 QTPNSKLVEYNPKSEDAAGAGETPHMLLRPHVMEPYETDMDLPRVFNQGIINPMLS 1047
DB 1081 qtpnslveynpksedagaggetphmlrrphvmeptyetdmdlprvfnqgiinpmis 1140
QY 1048 DETNICKGRKYVDGLTRNGTILVAFRGHYTWMLSPSPSPSAPARRITVWGIIPSDIVFT 1107
DB 1141 detnickgrkyvdgltrngtilvafrghytwmlspspspsaparritlewgiipspdivft 1200
QY 1108 RCNCEGKTFEPFKDSQYVRFNDIKDAGYPKPIFGEGGLGOIYAAALSTKRYKRWPSYV 1167
DB 1201 rcncegkttffkdsqywtfnndikdagypkpiifgglgoiyaaalstkykrmwpsyv 1260

QY 1158 FRFGSGIOQYIYKDEPVQKCGRRPALNYPYGEMTOVRRRRRPRATGPGSTHTIRIQ 1227
DB 1261 ftkrgsqlylykqepvqkcpgrpalnhyrgemtcvrrrrrrfrraagpsqthlrrlyq 1320
QY 1288 SPARLAYODKGLHNEVAVSILMRGLPNVVTSAISLPIRRKPDGYDYAFSKDOYNTIDV 1267
DB 1331 sparlalagdkglhnevavslmrqlpnvvtalslplrlrpdgydyafskdkqylnldv 1380
QY 1288 PSRRARATTTSGGTLSEKYNCP 1311
DB 1381 psrraralttrsggltskvyncp 1404

RESULT 2
AAB60568
ID AAB60568 standard; Protein: 1404 AA.
AC AAB60568;
XX
XX 27 APR-2001 (first entry)
DE Human megakaryocyte stimulating factor (MSF, CACP).
XX
XX Human: CACP protein; campodactylly-arthropathy-coxa vara-pericarditis;
KW MSF; megakaryocyte stimulating factor; synovial lubricant;
KM chromosome1q25-31; osteoarthritis; joint lubrication; osteopathic;
KW antirheumatic.
XX
XX Homo sapiens.
XX
XX WO200107068-A1.
XX
XX 01-FEB-2001.
XX
XX 21-JUL-2000; 2000MC-US20002.
XX
XX 23-JUL-1999; 9905-0145328.
XX 19-JUL-2000; 2000US-0145326.
XX
XX (UMCA-) UNIV CASE WESTERN RESERVE.
XX
XX Wadman ML;
XX
XX WPI: 2001-182721/18.
XX
XX New composition comprising the campodactylly-arthropathy-coxa
XX PT vara-pericarditis protein in combination with an anesthetic, useful for
XX PT treating osteoarthritis, or as lubricants of tissue and joints
XX
XX Example 1: Page -; 34pp: English
XX
XX The invention relates to a method of treating osteoarthritis via the
XX CC administration of a composition comprising the campodactylly-arthropathy-
XX CC coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.
XX CC The composition may further comprise a local anesthetic. The composition
XX CC of the invention may be administered via intra-articular or intravenous
XX CC injection. The human CACP protein is identified in the invention as
XX CC being megakaryocyte stimulating factor (MSF). The gene encoding
XX CC CACP protein (MSF) is located on chromosome 1q25-31, and mutations in
XX CC this gene are responsible for the heritable disorder campodactylly-
XX CC arthropathy-coxa vara-pericarditis, in which patients have synovial
XX CC hyperplasia without evidence of inflammation. CACP protein (MSF)
XX CC acts as a synovium lubricant, and can be used to lubricate tissue and
XX CC joints in the treatment of osteoarthritis. The composition may be
XX CC applied to reduce the symptoms of osteoarthritis (e.g., joint pain,
XX CC loss of range of movement or joint damage). The present sequence
XX CC represents human megakaryocyte stimulating factor (MSF, CACP protein).
XX CC Note: This sequence is not given in its entirety in figure 4 of the
XX CC specification, although a Genbank accession number was given. This
XX CC sequence was therefore obtained from Genbank (070316).
XX
XX Sequence 1404 AA:

Query Match 99.28; Score 7007.5; DB 22; Length 1404;
Best Local Similarity 93.4%; Pred. No. 0;
Matches 1311; Conservative 0; Mismatches 0; Indels 93; Gaps 1;
QY 1 1AAKTLPIYLLLSVFIQOYSSDQLSSCAGRCGGISRDATCNDYNCOHMECCPDF 60
DB 1 mwkltpllyllllsvfivqyvsqqlsscaagrcgyysrdctncdyngymecpdt 60
QY 61 KRYVTAELSCKRCGFESFERGECDDAOCRRYDKCCPPYESFCAE----- 106
DB 61 krcvdeelsckgrcfesfergrecddagckkydcpcpoyesfcaevhnpspsaktp 120
QY 107 ----- 106
DB 121 ppsagagqlsktckspkppnkkkkylseelteesvsengessssssssssstlw 180
QY 107 -----VKDNKKNRTKKRPKPPVDEGSLDNGDFRVTTPTST 147
DB 181 klksaknsaandqlqklkvdknknrtkkpvpvdaagsgldngdfkvtptdst 240
QY 148 TOHNVSTSPKINAKPINPESLPPNSDTSKETSIVNKEETVETKETTNNKOTSDG 207
DB 241 tqhmvstspkltakpnpvpslpnsdstaketslvnketvettettmkgustdg 300
QY 208 KEKTTSAKETQSIKETAADAPTSKYLAKPTPKAETTTKCPALTTPKEPTTPKPEAS 267
DB 301 kekttsaketqsietsadaptskylakpckpkaetltkypalttpepttkpeas 360
QY 268 TTPKEPTTTIKSAPTTPKEAPTTTSAPTTKEAPTTKEAPTTKEAPTTKEP 327
DB 361 ttpkeptttiksapttpkaptttsapttkeapttkeapttkeapttkeapttkeap 420
QY 328 APTTKSAPTTPKEPAPTTKRPAPTTPKAPTTPEPTTPPKEPAPTTKRP 387
DB 421 apttksapttpkapttkrapttpkapttpepttpkrapttkeapttkeapttkeap 480
QY 388 EPAPTAKKRPAPTTPKEPAPTTKRPAPTTKRPAPTTKRPAPTTKRPAPTT 447
DB 481 epaptakkrpapttkrapttkrapttkrapttkrapttkrapttkrapttkrap 540
QY 448 TTKSAPTTKEPSTTTKEAPTTKEAPTTKRPAPTTKRPAPTTKRPAPTTKRP 507
DB 541 ttksapttkepstttkeapttkeapttkrapttkrapttkrapttkrapttkrap 600
QY 508 APTAPKEPAPTTKEAPTTKRPAPTTKRPAPTTKRPAPTTKRPAPTTKRP 567
DB 601 aptapkepapttkrapttkrapttkrapttkrapttkrapttkrapttkrap 660
QY 568 PEAPATTTKRAAPNTPKEPAPTTKRPAPTTKRPAPTTKRPAPTTKRPAPTT 627
DB 661 peepattpkkaapntpkrapttkrapttkrapttkrapttkrapttkrapttkrap 720
QY 628 APTTPKRPAPKELAPTTKEPSTTSKAPPTPKGAPTTKEAPTTKEAPTTKEP 687
DB 721 apttpkrpapkelpkttkepsttskapttkgapttkgapttkgapttkgapttkg 780
QY 688 TAPTLKRPAPTTKRPAPKELAPTTKGPSTSDKRAPTTKRPAPTTKRPAPTT 747
DB 781 taptlkrappttkrpapkelpkttkgsptsdkrapttkrapttkrapttkrapttkr 840
QY 748 KPAPTTPPTPTTSEVSTPTTKEPTTIHNSPDESPETLSAETPKAENSKEGVP 807
DB 841 kpapttptpttsevspkttkepttihspsdespetlsaeptkalehspkegvp 900
QY 808 TKTPAATPEMTTAKKTEERDLFTPEPTTAAPOKTEATTTKTESKTTATTTQY 867
DB 901 tktpaatpemttaakkteerdlftpeptttaaoktetatttkteskttatttqy 960
QY 868 TSTTODTTPPKITTLTKTTLAPKVTYTTTKTITTEINMKPEETAKPKORATNSAK 927
DB 961 tstttdttpfkittltkttlapkvtttkkltteltimkpetakpkoratsakatck 1020

Mycobacterium tuberculosis
 Sequence gi|1011742
 Mycobacterium tuberculosis
 Sequence of the Pa
 Amino acid sequenc
 Amino acid sequenc
 Cryptosporidium pa
 protein encoded by
 Biobchative precu
 Portion of Cryptos
 Cryptosporidium pa
 Human OREF ORF995
 Human SRCP. Homo
 Human SRCP. Homo
 Human SRCP. Homo
 P. yoelii SS92 ant
 Peptide #2892 enco
 Peptide #2896 enco
 Human MUC11 polype
 C90P predicted am
 Human protein sequ
 Biobchative precu
 Peptide #1317 enco
 Peptide #1349 enco
 Peptide #1289 enco
 Human atrophin-1 r
 Human 07CG27 gene
 PRP 378. Triticum
 Streptococcus pneu
 Human OREF ORF2255
 PRP encoded by clo
 C. albicans Rbt1 p
 Peptide #967 enco
 Peptide #987 enco
 Peptide #941 enco

QY	140	KIKSSKNSAANLELOKKILVKNONKNKNTKKKPKPEPVDEAGSGLDNGPEKVTPTDST	199
Db	181	kikssksnaaneqlkklkvkdnkntkkpkpvyvdeaaagldngdfkvtcpbst	240
QY	200	TOHNVSVSPKTKTAKPILNPBPSLPBNSDSIKETSLIVNKEVETVETTTNNQSTDG	259
Db	241	tqhnvsvspklttkkpnprpslpnsdstketslvmkvetvettknqstdg	300
QY	260	KSKTSAKETOSIEKTSKADLAPISKYLAKTPRAEHTTGGPALTTPKEPTPTPKEPAS	319
Db	301	ksktsaeketsiektssakdlapiskylaktpraeettkgpaalttpkepttkepas	360
QY	320	TPPKBPPTTIKSAPTTPKEBAPTTPKSAAPTTPKEBAPTTPKBPAPTTPKBP	379
Db	361	tppkbppttlksaptpkbpaptpkbpaptpkbpaptpkbpaptpkbpaptpkbp	420
QY	380	APTTPKSAPTTPKEBAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBP	439
Db	421	aptpkbsaptpkbpaptpkbpaptpkbpaptpkbpaptpkbpaptpkbpaptpkbp	480
QY	440	EPAPAPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBP	499
Db	481	epapapkpaptpkbpaptpkbpaptpkbpaptpkbpaptpkbpaptpkbpaptpkbp	540
QY	500	TTKSAAPTTPKBPSPPTTKBAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBP	559
Db	541	ttksaptpkbpaptpkbpaptpkbpaptpkbpaptpkbpaptpkbpaptpkbpaptpkbp	600
QY	560	APTAPKBPAPTTPKETAAPTTPKKLPTTPBKLAPTTPKBPAPTTPBELAPTTPBEEPTPT	619
Db	601	aptapkeaptpcketapttpkkltpckpelahttpckpapttpelahttpckpapt	660
QY	620	PEEPAPTTPKAAAPNTPKBPAPTTPKBPAPTTPKBPAPTTPKETAAPTTPKGAAPTTLKEP	679
Db	661	peepaptpkaaapnckpkeaptpckpkeaptpckpkeaptpckpkeaptpckpkeaptpckp	720
QY	680	APTTPKPKBAPKELAPTTPKEBPTSTTSDKAPAPTTPKTAAPTTPKBPAPTTPKBPAPTTPKKG	739
Db	721	aptpkpapkelaptpkbpsttsdkapapttpktaapttpkbpaptpkbpaptpkkg	780
QY	740	TAPTTLKEBAPTTPKPKBAPKELAPTTPKBPSTTSDKAPAPTTPKETAAPTTPKBPAPTTPK	799
Db	781	tapttlkebaptpkpkbapkelaptpkbpsttsdkapapttpketahttpckpapt	840
QY	800	KPAPTTPPTPTSEVPTPTTKBETTHKSPDSTSELAEPTPKALBMSPKBPGVPT	859
Db	841	kpahttpptptsevppttkbetthkspdstselaeptpkalbmspkbpkgvpt	900
QY	860	TKTPAATKPBMTTAAKDKTTERDLKTPPETTTAAKMKETATTTETKEKIRATTTQV	919
Db	901	tktpaatkpbmtttaakdktterdlktppetttaakmketattttekksklatttqv	960
QY	920	TSTPTODTTPKRTITLAKTTTLAPKXTTAKITTTTLEINNKPEETAKPKDRATNSKATTPK	979
Db	961	tstptodttpkrtitlaktttlapkxttakittttleinnkpeetakpkdratnskattpk	1020
QY	980	POKPTAPKPKPSTIKKPKPMVBPVKPTTPPKMTSMPELNPTRSAEMLLOTTTRPN	1039
Db	1021	pokptapkpkpstikkpkpmvbpvkpttppkmtsmpeelnptrsaemllootttrpn	1080
QY	1040	QTPBNSKLIVEVNPBSDAGABEETPHMLRLRHVMPETVTPMDVLPRVPONGIITNPLS	1099
Db	1081	qtpbnsklivevnpbsdagabeetphmlrlrhvmpetvtpmdvlprvpngiitnplms	1140
QY	1100	DEMNICNGRPVNDGLTTLNGLTIVAEKGTIFMALSPEFSPBAPRRTIEWGJIPSEIDVFT	1159
Db	1141	demnicngrpvndglttlngltivaekgtifmalspefspbaprrtiewwgjipseidvft	1200
QY	1160	RCONCGKTEFFEDSOYMRFTNDIKDAGPKRIEFGFGLGQIYALSTAKYKNWPESVY	1219
Db	1201	rcnecgkttffedsgywrftndikdagypkriefggilcgqivaalstakynwpeasy	1260

QY 1210 FFKRGSIQOYIYKQEPYOKCPGRPALNYPVGEANTQVRRRPERAIGSPQHTIRIOY 1279
 Db 1261 ffxkxgslqylykqepvqkcpgrtrpnlmpvymqtvrrtrfetralsqctltiriy 1320
 QY 1280 SPALATYQDKGVLNEKXVSLMRGLPNVYTSALSPNIRKPDGYDYAASKQOYINIDV 1339
 Db 1321 sparllygdkgvlhnekvslwrglpvrvtsalslplrtkpdgydyatskdyndlv 1380
 QY 1340 PSTRARATRRSGOTLSRWYVNC 1363
 Db 1381 pstraratrrsgqlskvwnp 1404

RESULT 2

AA60568 standard; Protein: 1404 AA.

AA60568:

27-APR-2001 (first entry)

Human megakaryocyte stimulating factor (MSF, CACP).

Human; CACP protein; campodactyl-arthropathy-coxa vara-pericarditis;

MSF; megakaryocyte stimulating factor; synovial lubricant;

chromosome 12p25-31; osteoarthritis; joint lubrication; osteopathic;

antiarthritis.

Homo sapiens.

MO200107068-A1.

01-FEB-2001.

21-JUL-2000; 2000MO-US20002.

23-JUL-1999; 99US-045328.

19-JUL-2000; 2000US-0145328.

(DTCA-) UNIV CASE WESTERN RESERVE.

Warman ML;

WPI: 2001-182721/18.

New composition comprising the campodactyl-arthropathy-coxa

vara-pericarditis protein in combination with an anesthetic, useful for

treating osteoarthritis, or as lubricants of tissue and joints

Example 1: Page 7; 34pp; English.

The invention relates to a method of treating osteoarthritis via the
 administration of a composition comprising the campodactyl-arthropathy-
 coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.
 The composition may further comprise a local anesthetic. The composition
 of the invention may be administered via intra-articular or intravenous
 injection. The human CACP protein is identified in the invention as
 being megakaryocyte stimulating factor (MSF). The gene encoding
 CACP protein (MSF) is located on chromosome 12p25-31, and mutations in
 this gene are responsible for the heritable disorder campodactyl-
 arthropathy-coxa vara-pericarditis, in which patients have synovial
 hyperplasia without evidence of inflammation. CACP protein (MSF)
 acts as a synovial lubricant, and can be used to lubricate tissue and
 joints in the treatment of osteoarthritis. The composition may be
 applied to reduce the symptoms of osteoarthritis (e.g., joint pain,
 loss of range of movement or joint damage). The present sequence
 represents human megakaryocyte stimulating factor (MSF, CACP protein).
 Note: This sequence is not given in its entirety in figure 4 of the
 specification, although a Genbank accession number was given. This
 sequence was therefore obtained from Genbank (U07036).

Sequence 1404 AA;

Query Match 99.6%; Score 7245.5; DB 22; Length 1404;
 Best local Similarity 97.1%; Pred. No. 0;
 Matches 1363; Conservative 0; Mismatches 0; Indels 41; Gaps 1.

QY 1 MANKTPIYLLLLSVFYIQOYSSQ----- 25
 Db 1 manktpiylyllllsvfyiqvsgdlscagcggyrdctncdyncqhyamecpaf 60
 QY 26 -----ELSCNGRCEFESEFRGECODACQKYDCCCPYSEFCAEVNHPSPSSKAP 79
 Db 61 kvccchelsckgrcfesfergrecddagckkydkccpysesfcaevnmpspsskxap 120
 QY 80 PPGASQITKSTTKRSPKPNKKTKVLESEDEITEEHSVSENQSSSSSSSSSSSTIM 139
 Db 121 ppgasqitkstktrspkpnkkckkvleseeiteehsvsenqsssssssssstlw 180
 QY 140 KIKSSKNSANANDELQKLVKDNKNRTRKKYTPRPVYDEAGSLDNGDFVTPDST 199
 Db 181 kiksksnsaanrelqkllvkdnknrtkkytprrpvvdeags9ldngdfkvtcpdct 240
 QY 200 TQHNKYSTSPKATTAAPINPRLPNSDTSKETSLTVNKEETVETKETTNNKQSTNDG 259
 Db 241 tqhnkystspkltakpnlpslppnsdtskelsltnkettvetkettncqsltdg 300
 QY 260 KERTTSANETOSIEFTSAKDIAFTSKVLAKPPKAETTTGPAITPKREPTPKEPAS 319
 Db 301 kertsanetosieltksakdlaptskvlakppkaetttgpaaltprkptpkpas 360
 QY 320 TTPKREPTTTISAPTTAKREPAFTTTKSAPTTKREPAFTTTKREPAFTTTKEP 379
 Db 361 ttpkrepttllksapltckepapltksapltckepapltckepapltckepapltckep 420
 QY 380 APTTKSAPTTKREPAFTTTKREPAFTTTKREPAFTTTKREPAFTTTKREPAFTTK 439
 Db 421 apttksapttckepapltckepapltckepapltckepapltckepapltckepapltck 480
 QY 440 EPAPTPAKKPAFTTTKREPAFTTTKREPAFTTTKREPAFTTTKREPAFTTTKEP 499
 Db 481 epaptpakppapltckepapltckepapltckepapltckepapltckepapltckepap 540
 QY 500 TTKSAPTTKREPSPTTTKREPAFTTTKREPAFTTTKREPAFTTTKREPAFTTTKEP 559
 Db 541 tksapttckepsptttckepapltckepapltckepapltckepapltckepapltckep 600
 QY 560 APTAKREPAFTTTKREPAFTTTKREPAFTTTKREPAFTTTKREPAFTTTKREPAFT 619
 Db 601 aptakrepaftttckepapltckepapltckepapltckepapltckepapltckepaplt 660
 QY 620 PEEPAFTTPKAAAPNTKREPAFTTTKREPAFTTTKREPAFTTTKREPAFTTTKEP 679
 Db 661 peepafttpkaapntckepapltckepapltckepapltckepapltckepapltckepap 720
 QY 680 APTPKKAPKAPKELAPTTKREPTSTSDKAPTTKREPAFTTTKREPAFTTTKREPAFTTKG 739
 Db 721 aptpkkappkelpapltckepstsdapltckegtapltckepapltckepapltckepapltckg 780
 QY 740 TAPTLKREPAFTTTKREPAFTTTKREPAFTTTKREPAFTTTKREPAFTTTKREPAFTTK 799
 Db 781 taptlkrepafttpkpkpelpapltckegstsdapltckegtapltckepapltckepapltckp 840
 QY 800 KPAPTPPTPTPTSEVSTPTTTEPTTIHKSDEESTYELSAEPTPALNSPREPGVPT 859
 Db 841 kpaptptptptsevspttctkeptlhnksdesypelsaepcpalensprepgvpt 900
 QY 860 TKTPAATPEDETTTAXOKTTERDRLRTPENTTAAPKMTKEAATTTETKTSKITATTTQV 919
 Db 901 tktpaactpemttaaxoktlerdlrttpeettaapmktkeaattekttskitalttq 960
 QY 920 TSTTODTTPPKITTLKTTTLAPRVTTTKTITTTETLNNKPEEAKPKDRATNSKATTPK 979
 Db 961 tsttgdtpkltlktlktlaptvtttktttletlmmkpeeaakpkdratnaskattpk 1020

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:24:31 ; Search time 138.84 Seconds

(without alignments)
677.565 Million cell updates/sec

Title: AAS

Perfect score: 6814

Sequence: 1 MAWKTLPIYLLLSVFVQ.....ARAITRSQTSLSKWYNCP 1270

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_1101.*
1: /SID8/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID8/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID8/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID8/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID8/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID8/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID8/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID8/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID8/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID8/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID8/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID8/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID8/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID8/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID8/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID8/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID8/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID8/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID8/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID8/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID8/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Fred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6727	98.7	1404	13	MSF precursor. Sy
2	6727	98.7	1404	13	Human megakaryocyt
3	6727	98.7	1404	22	Human megakaryocyt
4	6151	90.3	1299	22	Human EST encoded
5	3484	51.1	902	22	Human MSF-derived
6	1707.5	25.1	472	22	Bovine MSF ortholo
7	1564	24.3	452	16	Human megakaryocyt
8	1171	17.2	5179	22	C899P predicted am
9	981	14.4	763	21	Arabidopsis thalia
10	950	13.9	1664	19	C. thermocellum OI
11	776.5	11.4	1325	22	Peptide #2327, enco

12	715	10.5	763	18	AAW31852	Mycobacterium tube
13	627	9.2	4412	21	AAV53666	Sequence g1/101742
14	572	9.0	572	18	AAV31855	Mycobacterium tube
15	547.5	8.0	844	7	AAE60570	Sequence of the fa
16	542	8.0	807	21	AAV54467	Amino acid sequenc
17	520.5	7.6	788	21	AAV54466	Amino acid sequenc
18	508	7.5	1837	21	AAV51726	Cryptosporidium pa
19	506.5	7.4	744	9	AAE82975	Bloodhesive precu
20	496.5	7.3	2971	21	AAV41231	Human ORF1 ORF95
21	489.5	7.2	2972	22	AAV50363	Human SRCP. Homo
22	489.5	7.2	3118	22	AAV50362	Portion of cryptos
23	489	7.2	1721	21	AAV11727	P. yoelii SP2 ant
24	488.5	7.2	826	13	AAV26042	Peptide #2869 enco
25	488	7.2	617	22	AAV16458	Human MUC11 poly
26	488	7.2	957	21	AAV59288	C900P predicted am
27	488	7.2	957	22	AAV24513	Cryptosporidium pa
28	488	7.2	1721	19	AAV48299	Human protein sequ
29	488	7.1	1127	22	AAV95341	Bloodhesive precu
30	485	7.0	652	9	AAE82974	Human SRCP. Homo
31	476.5	6.9	511	22	AAV14883	Portion of cryptos
32	467.5	6.9	511	22	AAV27312	Peptide #1317 enco
33	467.5	6.9	511	22	AAV27312	Peptide #1317 enco
34	467.5	6.9	511	22	AAV27312	Peptide #1289 enco
35	455.5	6.7	750	20	AAV05477	C. albicans Rb1 p
36	450.5	6.6	378	12	AAV14160	PRP 378. Trilicium
37	446.5	6.6	378	12	AAV14162	PRP encoded by clo
38	443	6.5	2870	21	AAV95559	Caenorhabditis ele
39	443	6.5	3178	21	AAV95559	Caenorhabditis ele
40	442.5	6.5	751	16	AAV80839	Japanese sea musse
41	437.5	6.4	2819	22	AAV35408	Human 07C627 gene
42	439	6.3	1510	21	AAV22870	EXFP-DEV0-MAP4-EBF
43	429	6.3	1510	21	AAV79648	Bifunctional caspa
44	428.5	6.3	2665	22	AAV14533	Peptide #967 enco
45	428.5	6.3	2665	22	AAV26950	Peptide #987 enco

ALIGNMENTS

RESULT 1	
ID	AAV26049 standard: Protein; 1404 AA.
XX	
AC	AAV26049;
DT	02-FEB-1993 (first entry)
XX	
DE	MSF precursor.
XX	
KW	Megakaryocyte colony stimulating factor; secretion signal; meg-CSF;
KW	stability; proteolytic cleavage; adhesion; alternative splicing.
OS	Synthetic.
XX	
FH	Key
FT	Region
FT	Location/Qualifiers
FT	1..26
FT	/label- Exon_I
FT	26..67
FT	/label- Exon_II
FT	67..107
FT	/label- Exon_III
FT	107..157
FT	/label- Exon_IV
FT	157..200
FT	/label- Exon_V
FT	200..1141
FT	/label- Exon_VI
FT	1411..1166
FT	/label- Exon_VII
FT	1166..1212
FT	/label- Exon_VIII
FT	1213..1266
FT	/label- Exon_IX

FT Region 1266..1331
 FT /label= Exon_X
 FT 1331..1373
 FT /label= Exon_XI
 FT 1373..1404
 FT /label= Exon_XII
 PN MO9213075-A.
 XX
 XX
 XX 06-AUG-1992.
 FD
 XX
 XX 17-JAN-1992; 92MO-US00433.
 PF
 XX 18-JAN-1991; 91US-0643502.
 PR 10-SEP-1991; 91US-0757022.
 PA (GENY) GENETICS INST INC.
 XX
 XX Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K;
 DR WP1: 1992-284660/34.
 DR N-PSDB: AA027223.
 XX
 FT New human mega-karyocyte stimulating factors - for treating
 PT immune deficiencies, cancer, exposure to radiation or drugs,
 PT bacterial and viral infections, etc.
 XX
 PS Claim 1, 2 and 3; Fig 1; 87pp; English.
 CC The sequence given is a full length translation from the megakaryocyte
 CC stimulating factor (MSF) precursor. The sequence covered by exons II,
 CC III and IV encodes megakaryocyte stimulating factor (MSF). This
 CC sequence is modified by the addition of an N-terminal sequence encoding
 CC a secretory leader, an initiating methionine proceeding exon II and a
 CC terminating codon following exon IV. The cDNA sequence given contains
 CC sequences derived from human megakaryocyte colony stimulating factor
 CC (meg-CSF). Exon I contains the initiating methionine, and encodes a
 CC classical mammalian protein secretion signal sequence. The sequence
 CC encoding the original meg-CSF includes exons II-IV and is thought to
 CC terminate in the region between amino acid residues 134 - 147. The
 CC primary transcript of this gene may be cleaved in different ways to
 CC yield a family of mRNAs each encoding a different MSF protein. Exons
 CC V and VI are thought to be related to the activity of the factor and
 CC are also implicated in the stability, folding and processing of the
 CC molecule. These exons are also thought to play a role in the observed
 CC synergy of MSF with other cytokines. Exons V - XII are believed to be
 CC implicated in the processing or folding of the appropriate structure of
 CC the resulting factor, ie. one or more of these exons may contain
 CC sequences which direct proteolytic cleavage, adhesion, organisation of
 CC the cellular matrix or extracellular matrix processing. Both naturally
 CC occurring and non-naturally occurring MSF's may be characterised by
 CC various combinations of alternatively spliced exons from this sequence,
 CC with the exons spliced together in differing orders to form different
 CC members of the MSF family.
 CC
 XX
 SQ Sequence 1404 AA:
 QY 1 MARKTPIYLLLSVFIQOVSSQ-----
 DB 1 maarktpiylililsvfiqvasqdlssacagrcgsgysrdatcncdyncqhyamecpdf 60
 QY 26 -----ELSCRGKCFSEFRGREGDCDDACCKRYDKCCPDYESFCAE----- 65
 DB 61 krtvctaelscgkrcfsefrgrecddacckkydkccpdyesfcsevhnpicspskkap 120
 QY 66 -----
 DB 121 ppsagatstktkrspkpkkkkkkvlseelteehtsvsengesssssssssstliw 180

QY 66 -----VKDNKNKNTKKKPPKPPVVDAGSGLDNGGFKVTPDPTST 106
 DB 181 kkesksaanrelqklkvlkndknkckkppkpvvdeagsgldngdfkvtlptst 240
 QY 107 TOHNKYSTSKITTAKEINRPSLPPNSDTSKETSILVNEKTVERTTTTKOSTDQ 166
 DB 241 tqhnkystskiltakpindrplppndtsketsilvnektvettlknqstldg 300
 QY 167 KEKTSKAKETOSIEKTSADIASTSVLAKPKPAETTTKGPALTTPKEEPTTPKEPAS 226
 DB 301 kektsaketqslektssakdlaptskvlakpkaetttkcpalttkpkepttkpas 360
 QY 227 TTKKEPTPTTISAPTPKEPAATTTKSAPTPKKEPAATTTKEPAATTTKEPAT 286
 DB 361 ttkkeptpttisaptpkepaatttksaptpkkepaatttkepaatttkepat 420
 QY 287 APTTKSAPTTKKEPAATTPPKPAATTPKEPAATTPKEEPTTPKEPAATTPKEPAT 346
 DB 421 apttksaptpkkepaattpkpaattpkepaattpkeeppttkkepat 480
 QY 347 EPAPTAPEKPAATTPKEPAATTPKEPAATTPKEEPTTPKEPAATTTKSAPT 406
 DB 481 epaptapekpaattpkepaattpkepaattpkeeppttkkepat 540
 QY 407 TTKSAPTPKKEPSPTTKEPAATTPKEPAATTPPKPAATTPKEPAATTPKEPAT 466
 DB 541 tksaptpkkespttkkepaattpkepaattpkpaattpkepaattpkepat 600
 QY 467 APAPKEPAATTPKEPAATTPPKKLTPTTPEKLAATTEKPAATTPKEEAPTPEEPT 526
 DB 601 apapekpaattpkepaattpkkltpttpeklaatttekpaattpkeeppt 660
 QY 527 PEEAPATTPPAAPNTPKEPAATTPKEPAATTPKEPAATTPKETAATTPGATPILKEP 586
 DB 661 peeapattpaapntpkepaattpkepaattpkepaattpketaattpgatlilkep 720
 QY 587 APTTPKPAKEKELAPTTKEPTSTSDKRAPTPKGTATPTTKEPAATTPKEPAT 646
 DB 721 apttpkpaekelapttkeptstsdkraptpkgtatpttkepaattpkepat 780
 QY 647 TAPTTLKEPAATTPKPAKEKELAPTTKGTSTSDKRAPTPKETAATTPKEPAT 706
 DB 781 tapttlkepaattpkpaekelapttkgtstsdkraptpketaattpkepat 840
 QY 707 KPAPTTPETPPPTTSVSPPTTKEPTTIHKSPPDESTEPLSAEPYPKALENPKRGVPT 766
 DB 841 kpapttpepppttsvsppttkcpttikhksppdestepelsaepypkalenskrgvpt 900
 QY 767 TTPAATKPEMTTAKDTEEDLRTTPEMTTAAKMKETAATTEKTESKITAATTTQV 826
 DB 901 tkpaatkpemttakdteedlrttpemttaakmketaattekteskitaattq 960
 QY 827 TSTTQODTTPRKITTLKTTTTLAPVYTTTKKTTTTEINMKPEETAKPDRATNSKATTPK 886
 DB 961 tsttqodttprkittllktttllapvytttkktttteinmkpeetakpdratnskattpk 1020
 QY 887 POKPTAKPKKPTSTKRPTRPRVRKPTTPPRKGTSTIMPELNPSTIAEAMQTTTTPN 946
 DB 1021 poptakpkkptstkrptprvrkpttpprkgtstimpelnpstiaeamqttttrpn 1080
 QY 947 QTPNSKLVENPKSEDAAGABETPHMLLRPHFMFPVTPDMQYLPRVPRNOGIIINPMS 1006
 DB 1081 qtpnslkvenpksedagabetphmlrrphfmfpvtpdmqylprvprnogiiinpms 1140
 QY 1007 DETNINCKGKPVVDGJTLTARNLTIAFRGHYFWMLSPSPSPARRITVEWGISPIDTVEF 1066
 DB 1141 detninckgkpvvdgjtltarnltiafrghyfwmlspspsparritvewgispidtvef 1200
 QY 1067 RCNCEGKTEFFFKDSQYWRFTNDIKDAGYPKPIFKGGGSLTGQIVAALESTAKKKNPESVY 1126
 DB 1201 rcncegktefffkdsqywrftndikdagypkpifkgggsltgqivaalestakknpsvy 1260

QY 1127 FFKRGSGTQOYIYKQEPVQKCPGRPALNYPVGEQVRRRRRPERAIGPSOHTIRIOY 1186
 Db 1161 ftkygsqylykqepqkcprrpalnypvgeqvrirrrrrerierisqscqtlrly 1320
 QY 1187 SPARLAVQDKGLHNEKVSILTMGLPNVYTSALSIPNIRKPDGYDYAFSKQOYTNIDY 1246
 Db 1121 sparlavqdgvlhnekvsltmglpnvvtalslplnrkpdgydyafskqdyndv 1380
 QY 1147 PSRRARATRRSGQTLKRWYVNC 1270
 Db 1181 psrrarattrsgqtlskwvynp 1404

RESULT 2

ID AAB60568 standard; Protein: 1404 AA.

AC AAB60568;

DE 27-APR-2001 (first entry)

Human megakaryocyte stimulating factor (MSF, CACP).

Human: CACP protein; campodactyl-arthropathy-coxa vara-pericarditis;
 MSF; megakaryocyte stimulating factor; synovial lubricant;
 chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;
 antirheptic.

OS Homo sapiens.

PN WC200107068-11.

PD 01-FEB-2001.

PF 21-JUL-2000; 2000OWO-US20002.

PR 23-JUL-1999; 99WS-0145328.

PR 19-JUL-2000; 2000OWO-0145328.

PA (DYCA-) UNIV CASE WESTERN RESERVE.

PI Wetman ML;

DR WEL; 2001-182721/18.

PT New composition comprising the campodactyl-arthropathy-coxa
 PT vara-pericarditis protein in combination with an anesthetic, useful for
 PS treating osteoarthritis, or as lubricants of tissue and joints
 PS Example 1; Page -: 34pp; English.

CC The invention relates to a method of treating osteoarthritis via the
 CC administration of a composition comprising the campodactyl-arthropathy-
 CC coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.
 CC The composition may further comprise a local anesthetic. The composition
 CC of the invention may be administered via intra-articular or intravenous
 CC injection. The human CACP protein is identified in the invention as
 CC being megakaryocyte stimulating factor (MSF). The gene encoding
 CC CACP protein (MSF) is located on chromosome 1q25-31, and mutations in
 CC this gene are responsible for the heritable disorder campodactyl-
 CC arthropathy-coxa vara-pericarditis, in which patients have synovial
 CC hyperplasia without evidence of inflammation. CACP protein (MSF)
 CC acts as a synovial lubricant, and can be used to lubricate tissue and
 CC joints in the treatment of osteoarthritis. The composition may be
 CC applied to reduce the symptoms of osteoarthritis (e.g., joint pain,
 CC loss of range of movement or joint damage). The present sequence
 CC represents human megakaryocyte stimulating factor (MSF, CACP protein).
 CC Note: This sequence is not given in its entirety in figure 4 of the
 CC specification, although a Genbank accession number was given. This
 CC sequence was therefore obtained from Genbank (U70316).

Sequence 1404 AA;

Query Match 98.7%; Score 6727; DB 22; Length 1404;
 Best Local Similarity 90.5%; Pred. No. 0;
 Matches 1270; Conservative 0; Mismatches 0; Indels 134; Gaps 2;

QY 1 MAWKTLPIYLLLSVEFIQVSSQ----- 25
 Db 1 mawktlpiylyllllsvfivqvsqdlsscagrgesyratdncdyncghmccpfd 60
 QY 26 -----ELSCGRGFESERERECDDAQCKKDYKCCPDYESFCAE----- 65
 Db 61 kvctaelscgrcfesfegrecdcagckkykccpdyesfcaevhnpssskkap 120
 QY 66 ----- 65
 Db 121 ppsasqdlkstkkrpkkpkkkklviseelteesvsenqssssssssatiw 180
 QY 66 -----VKDNKRNTRKKRPKPPVVDAGSGLDNGDFKTTPTST 106
 Db 181 klksksnaaanrelqkklkvkdnknrltkkpkpvpvdeasgldngfkvlcpdst 240
 QY 107 TOHNKYSTPKITTAAPINRPSLPRNSDSKETSLVNKEETVEKETTNNKOTSDG 166
 Db 241 tqhnkvstpsklttakpimrpslprnsdsketsltvnkettvettkqstidg 300
 QY 167 KEKTSKETOSEKTSANDLAPTSKVLAKPTPKAETTKGMLTTPKEPTTPKEPAS 226
 Db 301 kektsaketqsiketsadlaptskvlakpvpkaettkgmlttpkepttpkepas 360
 QY 227 TTPKEPTTPTKSAETTPKEPAETTKSAETTPKEPAETTPKEPAETTPKEPAETTK 286
 Db 361 ttpkepttptksaettpkepaettksaettpkepaettpkepaettpkepaettk 420
 QY 287 APTTKSAETTPKEPAETTPKEPAETTPKEPAETTPKEPAETTPKEPAETTPKE 346
 Db 421 apttksaettpkepaettpkepaettpkepaettpkepaettpkepaettpke 480
 QY 347 EPAPTRPKPAETTPKEPAETTPKEPAETTPKEPAETTPKEPAETTPKEPAET 406
 Db 481 epaptrpkpaettpkepaettpkepaettpkepaettpkepaettpkepaet 540
 QY 407 TTKSAETTPKEPSPTTKKEPAETTPKEPAETTPKEPAETTPKEPAETTPKEPA 466
 Db 541 ttksaettpkepspttkkepaettpkepaettpkepaettpkepaettpkepa 600
 QY 467 APTAKKEPAETTPKEPAETTPKEPAETTPKEPAETTPKEPAETTPKEPAET 526
 Db 601 aptakkepaettpkepaettpkepaettpkepaettpkepaettpkepaet 660
 QY 527 PEEPAETTPKAAETTPKEPAETTPKEPAETTPKEPAETTPKEPAETTPKEPAET 586
 Db 661 peepaettpkkaettpkepaettpkepaettpkepaettpkepaettpkepa 720
 QY 587 APTPKKPAKELAPTTKEPTSTSDKBAETTPKGAATTPKPAETTPKEPAETTP 646
 Db 721 aptpkkpaekelapttkeptstsdkbaettpkgaattpkpaettpkepaettp 780
 QY 647 TAPPTLEPAETTPKKAPELAPTTKGPSTSDKAPATTPKEPAETTPKEPAETTP 706
 Db 781 tapptlepaettpkkaapelapttkgpstsdkapattpkepaettpkepaettp 840
 QY 707 KPAETTPETPPTSEPTTKKEPTTHKSDSETPELASAPTRKALENSKEPGEVPT 766
 Db 841 kpaeptpetpptspttkkeptthksdsetpelasaptrkalenskepgevpt 900
 QY 767 TKPAAATKPEMTTAAKDTTERDLRTTPETTTAAPKGAETATTPKTESKITAT 826
 Db 901 tkpaaatkpemtttaakdttterdlrttpettttaapkgaeattpkteskitat 960
 QY 827 TSTTODTTPFKITLTKTTTAPKVTYTTTKTTTTEIMNKRPEETAKPKDRATNSKAT 886
 Db 961 tsttqdttpfkltlkttlapkvtytttkttttemnkrpeetakpkdratnsatk 1020

DR N-PSDB: AAR98981.
XX Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use
XX
PS Claim 20; Page 1198-1201; 1275pp; English.
XX
CC The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC biophysics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a
CC protein of the invention.
XX
SO Sequence 1299 AA

Query Match 100.0%; Score 5011; DB 22; Length 1299;
Best Local Similarity 100.0%; Pred. No. 1.5e-273;
Matches 941; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VQDNKRNKRRKPPKPPVDEAGSLNDGDKVTPDSTQHNKSTSPKTTAKPIN 60
DB 200 VQDNKRNKRRKPPKPPVDEAGSLNDGDKVTPDSTQHNKSTSPKTTAKPIN 259
QY 61 PRPSLPNSDTSKETSLSVNNKTTVEKETTNNKQSTDGKEKTSKESQISIKTSK 120
DB 260 PRPSLPNSDTSKETSLSVNNKTTVEKETTNNKQSTDGKEKTSKESQISIKTSK 319
QY 121 DLAPTSKTLAAPPKAEETTTGPAATTPKREPPTTPKPASTTPKREPPTTIKSAPTTPK 180
DB 220 DLAPTSKTLAAPPKAEETTTGPAATTPKREPPTTPKPASTTPKREPPTTIKSAPTTPK 379
QY 181 EPAPTTKSAATTPKREPPTTPKPAATTPKREPPTTIKSAATTPKREPPTTIKSA 240
DB 280 EPAPTTKSAATTPKREPPTTPKPAATTPKREPPTTIKSAATTPKREPPTTIKSA 439
QY 241 PKKPAATTPKREPPTTPKREPPTTIKSAATTPKREPPTTIKSAATTPKREPPTTIK 300
DB 440 PKKPAATTPKREPPTTPKREPPTTIKSAATTPKREPPTTIKSAATTPKREPPTTIK 499
QY 301 PTTTKEAPATTPKREPPTTPKREPPTTIKSAATTPKREPPTTIKSAATTPKREPPTTIK 360
DB 500 PTTTKEAPATTPKREPPTTPKREPPTTIKSAATTPKREPPTTIKSAATTPKREPPTTIK 559
QY 361 PAPTTPKREPPTTPKREPPTTIKSAATTPKREPPTTIKSAATTPKREPPTTIKSA 420
DB 560 PAPTTPKREPPTTPKREPPTTIKSAATTPKREPPTTIKSAATTPKREPPTTIKSA 619
QY 421 TPKKLTPTPEKLAATTPKREPPTTIKSAATTPKREPPTTIKSAATTPKREPPTTIK 480
DB 620 TPKKLTPTPEKLAATTPKREPPTTIKSAATTPKREPPTTIKSAATTPKREPPTTIK 679
QY 481 PAPTTPKREPPTTPKREPPTTIKSAATTPKREPPTTIKSAATTPKREPPTTIKSA 540
DB 680 PAPTTPKREPPTTPKREPPTTIKSAATTPKREPPTTIKSAATTPKREPPTTIKSA 739
QY 541 EPTSTSDKRAPATTPKGPATTPKREPPTTIKSAATTPKREPPTTIKSAATTPKREPPTTIK 600
DB 740 EPTSTSDKRAPATTPKGPATTPKREPPTTIKSAATTPKREPPTTIKSAATTPKREPPTTIK 799
QY 601 KELAPTTKGPSTSTSDKRAPATTPKGPATTPKREPPTTIKSAATTPKREPPTTIKSA 660
DB 800 KELAPTTKGPSTSTSDKRAPATTPKGPATTPKREPPTTIKSAATTPKREPPTTIKSA 859
QY 661 PTTTKEPTTIHKSPDSTPELSAFTPKALNSKKEGVPVTTTPATTKKEMTTAKDKT 720
DB 860 PTTTKEPTTIHKSPDSTPELSAFTPKALNSKKEGVPVTTTPATTKKEMTTAKDKT 919
QY 721 TERDLRTPEPTTAAPKMTKETATTEKTESKITATTTQVSTTODTTPKATITLTKTT 780
DB 920 TERDLRTPEPTTAAPKMTKETATTEKTESKITATTTQVSTTODTTPKATITLTKTT 979

DB 920 terdlrtpepttaapkncketattekteakitatqvtsttqdtltpfkittlkt 979
QY 761 TLAPKVTTKKRTTTTTEIANNPEETAKPRDRATNSKATTPKPKPTKPKSTKPKPT 840
DB 980 TLAPKVTTKKRTTTTTEIANNPEETAKPRDRATNSKATTPKPKPTKPKSTKPKPT 1039
QY 841 MPRVAKPTTPPKKSTWPELANPNSRIABAMLOTTTRPQNPNSKLVEVNPKEADAG 900
DB 1040 mprvaxkpttpkkskswpeelanpnsriaabamlotttrpnpnsklvevnpkedsag 1099
QY 901 AEGEPHLLLRPHVMPKPRVNDMDYLRVNPQGIINMWS 941
DB 1100 aegephlllrphvmpkprvndmdylrvnpqgillnmpws 1140

RESULT 2
AAR26049
ID AAR26049 standard; Protein: 1404 AA.
AC AAR26049;
XX
DT 02-FEB-1993 (first entry)
XX
DE MSF precursor.
XX
KM Megakaryocyte colony stimulating factor; secretion signal; meg-CSF;
KM stability; proteolytic cleavage; adhesion; alternative splicing.
XX
OS Synthetic.

XX
FH Key Location/Qualifiers
FT Region 1..26 /label= Exon_I
FT Region 26..67 /label= Exon_II
FT Region 67..107 /label= Exon_III
FT Region 107..157 /label= Exon_IV
FT Region 157..200 /label= Exon_V
FT Region 200..1141 /label= Exon_VI
FT Region 1141..1166 /label= Exon_VII
FT Region 1166..1212 /label= Exon_VIII
FT Region 1212..1266 /label= Exon_IX
FT Region 1266..1331 /label= Exon_X
FT Region 1331..1373 /label= Exon_XI
FT Region 1373..1404 /label= Exon_XII
FT Region 1404..1404 /label= Exon_XII

MO9213075-A.
PD 06-AUG-1992.
XX
PF 17-JAN-1992; 92MO-US00433.
XX
PR 18-JAN-1991; 91US-0643502.
PR 10-SEP-1991; 91US-0757022.
XX
PA (GENE) GENETICS INST INC.
XX
PI Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K;
XX WPI; 1992-284660/34.
DR N-PSDB; AAO27223.
XX
PT New human mega-karyocyte stimulating factors - for treating

PT Immune deficiencies, cancer, exposure to radiation or drugs,
 PT bacterial and viral infections, etc.

XX Cladm 1, 2 and 3; Fig 1; 87pp; English.

CC The sequence given is a full length translation from the megakaryocyte
 CC stimulating factor (MSF) precursor. The sequence covered by exons II,
 CC III and IV encodes megakaryocyte stimulating factor (MSF). This
 CC sequence is modified by the addition of an N-terminal sequence encoding
 CC a secretory leader, an initiating methionine preceding exon II and a
 CC terminating codon following exon IV. The cDNA sequence given contains
 CC sequences derived from human megakaryocyte colony stimulating factor
 CC (meg-CSF). Exon I contains the initiating methionine, and encodes a
 CC classical mammalian protein secretion signal sequence. The sequence
 CC encoding the original meg-CSF includes exons II-IV and is thought to
 CC terminate in the region between amino acid residues 134 - 147. The
 CC primary transcript of this gene may be cleaved in different ways to
 CC yield a family of mRNA's each encoding a different MSF protein. Exons
 CC V and VI are thought to be related to the activity of the factor and
 CC are also implicated in the stability, folding and processing of the
 CC molecule. These exons are also thought to play a role in the observed
 CC synergy of MSF with other cytokines. Exons V - XII are believed to be
 CC implicated in the processing or folding of the appropriate structure of
 CC the resulting factor, i.e. one or more of these exons may contain
 CC sequences which direct proteolytic cleavage, adhesion, organisation of
 CC the cellular matrix or extracellular matrix processing. Both naturally
 CC occurring and non-naturally occurring MSF's may be characterised by
 CC various combinations of alternatively spliced exons from this sequence,
 CC with the exons spliced together in differing orders to form different
 CC members of the MSF family.

XX Sequence 1404 AA:

Query Match 100.0%; Score 5011; DB 13; Length 1404;
 Best Local Similarity 100.0%; Pred. No. 1,6e-273;
 Matches 941; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VNDNKNRTKRRKPPVNDVDEAGSLDNGDFVTTTDDSTTQHNKVSPPKTKAPIN 60
 DB 200 VKDNNKRRKRRKPPVNDVDEAGSLDNGDFVTTTDDSTTQHNKVSPPKTKAPIN 259
 QY 61 PRPSLPNDSTKETSIVNKKETVETKETTNNKQSTDGKETSKEQSISAK 120
 DB 260 PRPSLPNDSTKETSIVNKKETVETKETTNNKQSTDGKETSKEQSISAK 319
 QY 121 DLAPTSKVLAKRPPKAEETTKGPALETTKPEPTTTPKBPASTTKSAPTTK 180
 DB 320 DLAPTSKVLAKRPPKAEETTKGPALETTKPEPTTTPKBPASTTKSAPTTK 379
 QY 181 EPAFTTTSAPTTKPEAPTTTKEBPATTTKBPATTTKBPATTTKBPATTT 240
 DB 380 EPAFTTTSAPTTKPEAPTTTKEBPATTTKBPATTTKBPATTTKBPATTT 439
 QY 241 PKRPAPTTKPEAPTTTKEBPATTTTKEBPATTTTKEBPATTTTKEBPAT 300
 DB 440 PKRPAPTTKPEAPTTTKEBPATTTTKEBPATTTTKEBPATTTTKEBPAT 499
 QY 361 PTPPEAPTTTKEBPATTTTKEBPATTTTKEBPATTTTKEBPATTTTKEBP 360
 DB 500 PTPPEAPTTTKEBPATTTTKEBPATTTTKEBPATTTTKEBPATTTTKEBP 559
 QY 361 PAPTTPKRPAPTTTKEBPATTTTKEBPATTTTKEBPATTTTKEBPATTT 420
 DB 560 PAPTTPKRPAPTTTKEBPATTTTKEBPATTTTKEBPATTTTKEBPATTT 619
 QY 421 TPKKLTPTTPEKLAPTTPEKAPTTPEELAPTTPEEPPTTPEEPPTTPEEP 480
 DB 640 TPKKLTPTTPEKLAPTTPEKAPTTPEELAPTTPEEPPTTPEEPPTTPEEP 679
 QY 481 PAPTTPKRPAPTTTKEBPATTTTKEBPATTTTKEBPATTTTKEBPATTT 540
 DB 680 PAPTTPKRPAPTTTKEBPATTTTKEBPATTTTKEBPATTTTKEBPATTT 739

QY 541 EPTSTSDKPAPTTPKGAAPTTPKBPATTTKBPATTTKBPATTTKBPATTT 600
 DB 740 EPTSTSDKPAPTTPKGAAPTTPKBPATTTKBPATTTKBPATTTKBPATTT 799
 QY 601 KELAPTTKGPSTTSDKPAPTTPKBPATTTKBPATTTKBPATTTKBPATTT 660
 DB 800 KELAPTTKGPSTTSDKPAPTTPKBPATTTKBPATTTKBPATTTKBPATTT 859
 QY 661 PTTTEPTTTRKSPESPEPELSAETTPALNSPKREVPTTKTPAATKPMATTK 720
 DB 860 PTTTEPTTTRKSPESPEPELSAETTPALNSPKREVPTTKTPAATKPMATTK 919
 QY 721 TERDRTTPTTAAFKTKETATTTKTSKITATTOYSTTTOOTPTTKTTTAKT 780
 DB 920 TERDRTTPTTAAFKTKETATTTKTSKITATTOYSTTTOOTPTTKTTTAKT 979
 QY 781 TLAPVTTTKKTTTTEITMNRPEETAKRKDRATNSKATTPPKPKTKPKTKPK 840
 DB 980 TLAPVTTTKKTTTTEITMNRPEETAKRKDRATNSKATTPPKPKTKPKTKPK 1039
 QY 841 MPVAKPPTTPTPKMTSTMPLELNPSTRIANMLOTTTRPQNTNSKLVENPKSEDAG 900
 DB 1040 MPVAKPPTTPTPKMTSTMPLELNPSTRIANMLOTTTRPQNTNSKLVENPKSEDAG 1099
 QY 901 AEGETPHMLRPHVFMPEVTPDMOYLPRVNOGIINPMLS 941
 DB 1100 AEGETPHMLRPHVFMPEVTPDMOYLPRVNOGIINPMLS 1140

RESULT 3

AAB60568 AAB60568 standard; Protein; 1404 AA.

AC AAB60568;

DT 27 APR-2001 (first entry)

DE Human megakaryocyte stimulating factor (MSF, CACP).

KM Human; CACP protein; camploclactyly-arthropathy-coxa vara-pericarditis;

KW MSF; megakaryocyte stimulating factor; synovial lubricant;

KW chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;

OS antidiabetic.

OS Homo sapiens.

PF 21-JUL-2000; 2000OMO-US20002.

PR 23-JUL-1999; 99US-0145328.

PR 19-JUL-2000; 2000US-0145328.

PA (UYCA-) UNIV CASE WESTERN RESERVE.

PI Warman ML;

DR WPI; 2001-182721/18.

PT New composition comprising the camploclactyly-arthropathy-coxa

PT var-pericarditis protein in combination with an anesthetic, useful for

PT treating osteoarthritis, or as lubricants of tissue and joints

PS Example 1; Page 34pp; English

CC The invention relates to a method of treating osteoarthritis via the

CC administration of a composition comprising the camploclactyly-arthropathy-

CC coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.

CC The composition may further comprise a local anesthetic. The composition

CC of the invention may be administered via intra-articular or intravenous

SEA ID NO: 1, 200-1167
AC NO: AAR26049, A-gensay-1101
File Copy
Page 219

DR N-PSDB: AAR98981.

Isolated polypeptide for treatment of diseases, diagnostics, raising antibodies and research use

Claim 20: Page 1198-1201: 1275pp; English.

The present invention provides the protein and coding sequences of novel proteins from a variety of organisms, including human, dog, cat, horse, cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea urchin and tomato. These were derived from expressed sequence tags (ESTs) from the organism of interest. They can be used in diagnostics, forensics, gene mapping, identification of mutations, to assess biodiversity and for nutritional purposes. The present sequence is a protein of the invention.

Sequence 1299 AA:

Query Match 100.0%; Score 5155; DB 22; Length 1299;
Best Local Similarity 100.0%; Pred. No. 3.9e-282;
Matches 968; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VKDNKKNTKPKPKPVVDAGSGLDNGDFKVTPTSTTOHNVSTSPKITTAKPIN 60
200 vkdnkkntkpkpkpvvdagsgldngdfkvtptsttohnvstspkittakpin 259
61 PRSLPFRNSDSKESLTVNKETTVEKETTNNKOSTDGEKETSARETOSIEKTSAK 120
260 prslpfrnsdskesltvnkettvetkettnnkostdgeketsaretosiekttsak 319
121 DLAPTSKILAKPKAPATTGGPALTTPKEPTTPPKKPAATTPKKEPTTTPKSAPTTPK 180
320 dlaptskilakpkapattggpalttpkepttpkkpaattpkkekptttpksapttpk 379
181 EPAPTTKSAATTPKKEPAATTPKKEPAATTPKKEPAATTPKKEPAATTPKKEPAATTP 240
380 epapttksaattpkkepaattpkkepaattpkkepaattpkkepaattpkkepaattp 439
241 PKRAPTPKKEPAATTPKKEPAATTPKKEPAATTPKKEPAATTPKKEPAATTPKKEPA 300
440 pkraptpkkepaattpkkepaattpkkepaattpkkepaattpkkepaattpkkepa 499
301 PTPKKEPAATTPKKEPAATTPKKEPAATTPKKEPAATTPKKEPAATTPKKEPAATTP 360
500 ptpkkepaattpkkepaattpkkepaattpkkepaattpkkepaattpkkepaattp 559
361 PAPTTPKKEPAATTPKKEPAATTPKKEPAATTPKKEPAATTPKKEPAATTPKKEPAAT 420
560 papttpkkepaattpkkepaattpkkepaattpkkepaattpkkepaattpkkepaat 619
421 TPKKLTPTPEKLAATTPKKEPAATTPKKEPAATTPKKEPAATTPKKEPAATTPKKEPA 480
620 tpkkltptpeklaattpkkepaattpkkepaattpkkepaattpkkepaattpkkepa 679
481 PAPTTPKKEPAATTPKKEPAATTPKKEPAATTPKKEPAATTPKKEPAATTPKKEPAATTP 540
680 papttpkkepaattpkkepaattpkkepaattpkkepaattpkkepaattpkkepaat 739
541 EPTSTSDKAPATTPKKEPAATTPKKEPAATTPKKEPAATTPKKEPAATTPKKEPAATTP 600
740 eptstsdkapattpkkepaattpkkepaattpkkepaattpkkepaattpkkepaat 799
601 KELAPTTKGPSTSTSKAPATTPKKEPAATTPKKEPAATTPKKEPAATTPKKEPAATTP 660
800 kelapttkgpststskapattpkkepaattpkkepaattpkkepaattpkkepaattp 859
861 PTTTKEPTTHKSPDESPKSAEPKALENSKEPVTTPKPAATTPKKEPAATTPKKEPAATTP 720
860 ptttkeptthkspdespksaepkalenskepvttpkpaattpkkepaattpkkepaattp 919
721 TERDLATPTETTTAARMKKETAATTEKTESKITATTTTSTTODTTPPKITLTKTT 780
|||||

Db 920 terdlatptetttaarmlketaattekteskitatTTTSTTODTTPPKITLTKTT 979
QY 781 TLAPKVTNTKTTTTEIMNKEETANKPRDRATNSKATTPKPKPKPTSKPKPT 840
Db 980 tlapkvtntktttteimnkeetankprdratnskattpkpkpkptskpkpt 1039
QY 841 MPRVRKRTTPPKKNTSTNELNPSTRIABAMLOTTTRPNOTPNSKIVEVNPSEADAG 900
Db 1040 mprvrkrttppkkntstnelnpstriaabamloTTTRPNOTPNSKIVEVNPSEADAG 1099
QY 901 AEGTPHMLRPHVEMPEVPMOYLPRVPMOGIITNPMUSDENICNGKRVDDGTTLRN 960
Db 1100 aegtpHMLRPHVEMPEVPMOYLPRVPMOGIITNPMUSDENICNGKRVDDGTTLRN 1159
QY 961 GTLVAFRG 968
Db 1160 gltvafrg 1167

RESULT 2
AAR26049
ID AAR26049 standard; Protein: 1404 AA.
XX
AC AAR26049;
XX
DT 02-FEB-1993 (first entry)
XX
DE MSF precursor.
XX
KM Megakaryocyte colony stimulating factor; secretion signal; meg-CSF;
XX stability; proteolytic cleavage; adhesion; alternative splicing.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Region 1..26 /label= Exon_I
FT Region 26..67 /label= Exon_II
FT Region 67..107 /label= Exon_III
FT Region 107..157 /label= Exon_IV
FT Region 157..200 /label= Exon_V
FT Region 200..1141 /label= Exon_VI
FT Region 1411..1166 /label= Exon_VII
FT Region 1166..1212 /label= Exon_VIII
FT Region 1213..1266 /label= Exon_IX
FT Region 1266..1331 /label= Exon_X
FT Region 1331..1373 /label= Exon_XI
FT Region 1373..1404 /label= Exon_XII
PN MO9213075-A.
XX
PD 06-AUG-1992.
XX
PF 17-JAN-1992; 92WO-US00433.
XX
PR 18-JAN-1991; 91US-0643502.
XX 10-SEP-1991; 91US-0757022.
XX (GENY) GENETICS INSTR INC.
XX Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K;

DR WPI: 1992-284660/34.
 DR N-ESDB: AAO27223.
 PT New human mega-karyocyte stimulating factors - for treating
 PT immune deficiencies, cancer, exposure to radiation or drugs,
 PT bacterial and viral infections, etc.
 PS C14im 1, 2 and 3; Fig 1; 87pp; English.

CC The sequence given is a full length translation from the megakaryocyte
 CC stimulating factor (MSF) precursor. The sequence covered by exons II,
 CC III and IV encodes megakaryocyte stimulating factor (MSF). This
 CC sequence is modified by the addition of an N-terminal sequence encoding
 CC a secretory leader, an initiating methionine preceding exon II and a
 CC terminating codon following exon IV. The cDNA sequence given contains
 CC sequences derived from human megakaryocyte colony stimulating factor
 CC (meg-CSF). Exon I contains the initiating methionine, and encodes a
 CC classical mammalian protein secretion signal sequence. The sequence
 CC encoding the original meg-CSF includes exons II-IV and is thought to
 CC terminate in the region between amino acid residues 134 - 147. The
 CC primary transcript of this gene may be cleaved in different ways to
 CC yield a family of mRNA's each encoding a different MSF protein. Exons
 CC V and VI are thought to be related to the activity of the factor and
 CC are also implicated in the stability, folding and processing of the
 CC molecule. These exons are also thought to play a role in the observed
 CC synergy of MSF with other cytokines. Exons V - XII are believed to be
 CC implicated in the processing or folding of the appropriate structure of
 CC the resulting factor. ie one or more of these exons may contain
 CC sequences which direct proteolytic cleavage, adhesion, organisation of
 CC the cellular matrix or extracellular matrix processing. Both naturally
 CC occurring and non-naturally occurring MSF's may be characterised by
 CC various combinations of alternatively spliced exons from this sequence,
 CC with the exons spliced together in differing orders to form different
 CC members of the MSF family.

Sequence 1404 AA:

Query Match 100.0%; Score 5155; DB 13; Length 1404;
 Best Local Similarity 100.0%; Pred. No. 4.2e-282;
 Matches 968; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDNKKRRTKRRKPPKPPVDEAGSLDNGDFKVTTPDTSTTQHNKVSPTTKAKPIN 60
 DB 200 vkdknkrrtkkpkpvpvdeagsgldngdfkvttpdsttqhnkvspttkakpin 259
 QY 61 PRSLPNSDTSKETSLSYTKNKTETTTTNTKOTSTDGKTKTSAKTSQSIKTSK 120
 DB 260 ppslpnsdtsketsalsynkeltvktcttnkqtscdgktsaktsqsietsak 319
 QY 121 DLAPTSKVLAKPPKAETTTKGPALTTPKEPTPTTPKPEASTTPKEPTTTKSAPTPK 180
 DB 320 dlaptskvllakppkaetttkgtalttpeptpttpkpeasttpkeptttksaptcpk 379
 QY 181 EAPATTTKSAPTTPKPAATTTKPAATTTKPAATTTKPAATTTKSAPTTPKPAAT 240
 DB 380 eapatttksapttppkpaatttkpaatttkpaatttkpaatttksapttcpkpaapt 439
 QY 241 PKKPAATTTKPEAPPTTPKPEPTPTTPKPEAPTTKPEAPTTKPAKPAATTPKPE 300
 DB 440 pkkpatttkpeappttpkpeptpttpkpeapttkpeapttkpaakpaattpkpe 499
 QY 301 PTTTPKPEAPTTKPEAPTTKPEAPTTKSAPTTTKSAPTTTKSAPTTPKPEAPT 360
 DB 500 ptttpkpeapttkpeapttkpeapttksaptttkpeapttksaptcpkpeapttke 559
 QY 361 PAPTTPKPEAPTTKPAATTTKPEAPTTKPAATTTKPAATTTKPAATTPKETAAPT 420
 DB 560 papttpkpeapttkpaatttkpeapttkpaatttkpaatttkpaaptcpketaapt 619
 QY 481 TPKKLLPTTPKELAPTTPEKPAATTPPEELAPTTPEEPTPTTPKEEAPPTTPKAA 480
 DB 680 tpkkllpttpkelapttpekpaattppeelapttpeeptpttpkeeappttpkkaa 679

QY 481 PAPTTPKPEAPTTKPEAPTTTPKETAAPTTPKGTAPTTKPEAPTTPKKPAKELAPTTK 540
 DB 680 papttpkpeapttkpeaptttpketaapttpkgtapttkpeapttpkkpaakelaptttk 739
 QY 541 EPTSTTSQKPAATTTKGAAPTTKPEAPTTKPEAPTTKGAAPTTKPEAPTTKPKAP 600
 DB 740 eptsttsdkpaatttkgaapttkpeapttkpeapttkgaapttkpeapttkpkap 799
 QY 601 KELAPTTKGPSTSDKPAATTPKETAAPTTPKPEAPTTKPKKPAATTPETPTTSYST 660
 DB 800 kelapttkgsstsdkpaattpketaapttpkpeapttkpkkpaattpetpttsyst 859
 QY 661 PTTTKEPTTIHKSPPESSTPELSAEPPTKALNSKPEKCVPTTKTPAATKPEATTAKKT 720
 DB 860 ptttkepttihsppesstpelasaepptkalnskpekevpttktpaatkpeattakkt 919
 QY 721 TPRLDRTPTPTTAAKPKTKETATTTETTSKRTATTTQVSTTTQDTPTFKITTLKTT 780
 DB 920 terldrtptptttaaakpktketatttetskrtatttqvtstttqdtptfkittlkt 979
 QY 781 TLAPVTTTKKTTTTEIMNKPEETAKPKDRATNSKATTPKOKTPAKPKPTSKRKT 840
 DB 980 tlavtttkktttteimnkpeetakpkdratnskattpoktpakpkptskrkkt 1039
 QY 841 MPRVRKPTTPTPPKMTSTMBELNPTSRISAEAMLTTRPNQTNNSKLVENPKSEDAAG 900
 DB 1040 mprvrkpttptppkmtstmbelnptsrisaeamlttrpntqtnnsklvenpksedagg 1099
 QY 901 AEGEPHMLLPHYFMPTVTDNMYLPRVNOGIIINPMLSDEFINICGRVYDGLTTLRN 960
 DB 1100 aegephmlllphymptvtdnmylprvnoqiinpmldsdefinigrvydglttlrn 1159
 QY 961 GTLVAFRG 968
 DB 1160 gtlvafrg 1167

RESULT 3
 ID AAB6568 standard; Protein; 1404 AA.
 AC AAB60568;
 DT 27-APR-2001 (first entry)

DE Human megakaryocyte stimulating factor (MSF, CACP).
 DE Human megakaryocyte stimulating factor (MSF, CACP).
 KW MSF; megakaryocyte stimulating factor; synovial lubricant;
 KW chromosome 1q23.31; osteoarthritis; joint lubrication; osteopathic;
 KW antiarthritic.
 OS Homo sapiens.
 PN WO200107068-A1.
 PD 01-FEB-2001.
 PE 21-JUL-2000; 2000MO-US20008.
 PR 23-JUL-1999; 99US-0145328.
 PR 19-JUL-2000; 2000US-0145328.
 PA (UYCA-) UNIV CASE WESTERN RESERVE.

Warman ML;
 WPI: 2001-182721/18.

New composition comprising the camptodactylly-arthropathy-coxa
 vara-pericarditis protein in combination with an anesthetic, useful for
 treating osteoarthritis, or as lubricants of tissue and joints.

SE 1D NO: 1, 200-1212
AC NO: AAR26049, A-GeneSeq-1104
File copy
Page 2

NRPSDB: AAR98981.

Isolated polypeptide for treatment of diseases, diagnostics, raising antibodies and research use.

Clalim 20: Page 1198-1201; 1275pp; English.

The present invention provides the protein and coding sequences of novel proteins from a variety of organisms, including human, dog, cat, horse, cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea urchin and tomato. These were derived from expressed sequence tags (ESTs) from the organism of interest. They can be used in diagnostics, forensics, gene mapping, identification of mutations, to assess biodiversity and for nutritional purposes. The present sequence is a protein of the invention.

Sequence 1299 AA:

Query Match 100.0%; Score 5416; DB 22; Length 1299;
Best Local Similarity 100.0%; Pred. NO. 2.4e-298;
Matches 1013; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VKDNKKNTKKKPTKPPVNDAGSGLDNGDFKVTTPDTSTTOHNVKSTSPKITTAKPIN 60
200 VKDNKKNTKKKPTKPPVNDAGSGLDNGDFKVTTPDTSTTOHNVKSTSPKITTAKPIN 259
61 PRPSLPKRSQTSKETSIVNKKETVERKETTNNKQSTDSGEKTTSAKQSTIEKTSAK 120
260 PRPSLPKRSQTSKETSIVNKKETVERKETTNNKQSTDSGEKTTSAKQSTIEKTSAK 319
121 DLASTSKVLAAPKPAETTTKAPLTPPKPTTPPKPTTPPKPTTPPKPTTPPKPTTPK 180
320 dlaptskvlakpkaetttkapalettpkpttpkpttpkpttpkpttpkpttpkpttpk 379
181 EPAATTTKAPTTKREAPATTTKAPATTKREAPATTKREAPATTKREAPATTKREAPATT 240
380 epaatTTKAPTTKREAPATTTKAPATTKREAPATTKREAPATTKREAPATTKREAPATT 439
241 PKRAPATTPKREAPATTTPKREAPATTTPKREAPATTTPKREAPATTTPKREAPATT 300
440 PKRAPATTPKREAPATTTPKREAPATTTPKREAPATTTPKREAPATTTPKREAPATT 499
301 PTPKREAPATTTPKREAPATTTPKREAPATTTPKREAPATTTPKREAPATTTPKREAP 360
500 PTPKREAPATTTPKREAPATTTPKREAPATTTPKREAPATTTPKREAPATTTPKREAP 559
361 PAPTTPKREAPATTTPKREAPATTTPKREAPATTTPKREAPATTTPKREAPATTTPK 420
560 PAPTTPKREAPATTTPKREAPATTTPKREAPATTTPKREAPATTTPKREAPATTTPK 619
421 TPKKLTTPTPKREAPATTTPKREAPATTTPKREAPATTTPKREAPATTTPKREAPATT 480
620 TPKKLTTPTPKREAPATTTPKREAPATTTPKREAPATTTPKREAPATTTPKREAPATT 679
481 PAPTTPKREAPATTTPKREAPATTTPKREAPATTTPKREAPATTTPKREAPATTTPK 540
680 PAPTTPKREAPATTTPKREAPATTTPKREAPATTTPKREAPATTTPKREAPATTTPK 739
541 EPNTSTSDKAPATTPKREAPATTTPKREAPATTTPKREAPATTTPKREAPATTTPK 600
740 EPNTSTSDKAPATTPKREAPATTTPKREAPATTTPKREAPATTTPKREAPATTTPK 799
601 KELAPATTPKREAPATTTPKREAPATTTPKREAPATTTPKREAPATTTPKREAPATT 660
800 KELAPATTPKREAPATTTPKREAPATTTPKREAPATTTPKREAPATTTPKREAPATT 859
661 PTTTKEPTTHKSPDESTPELSAEPTRKALENSKPEGVPTTKTPATKREMTTAKDKT 720
860 PTTTKEPTTHKSPDESTPELSAEPTRKALENSKPEGVPTTKTPATKREMTTAKDKT 919
721 TEBDLRTTPETTTAARPMETATTTTETTESKLTATTTTVOVSTTTTODTTPKITTAKT 780

Db 920 TELRTTPETTTAARPMETATTTTETTESKLTATTTTVOVSTTTTODTTPKITTAKT 979
QY 781 TLAPRTTTPKRTTTPKRTTTPKRTTTPKRTTTPKRTTTPKRTTTPKRTTTPKRTT 840
Db 980 TLAPRTTTPKRTTTPKRTTTPKRTTTPKRTTTPKRTTTPKRTTTPKRTTTPKRTT 1039
QY 841 MPVRKRTTTPKRTTTPKRTTTPKRTTTPKRTTTPKRTTTPKRTTTPKRTTTPKRTT 900
Db 1040 MPVRKRTTTPKRTTTPKRTTTPKRTTTPKRTTTPKRTTTPKRTTTPKRTTTPKRTT 1099
QY 901 AEGTTPMLRPRVPMEMVPRDDVLPVRVNOGIIINPMISDSTNINCGKPDVGLITLRN 960
Db 1100 AEGTTPMLRPRVPMEMVPRDDVLPVRVNOGIIINPMISDSTNINCGKPDVGLITLRN 1159
QY 961 GTVAERGHYFMMLSPPSPSPAPRTTEVWGISPIDVTFRNCDEGKTFEFK 1013
Db 1160 GTVAERGHYFMMLSPPSPSPAPRTTEVWGISPIDVTFRNCDEGKTFEFK 1212

RESULT 2

AAR26049 standard; Protein; 1404 AA.

AAR26049;

02-FEB-1993 (first entry)

MSF precursor.

Megakaryocyte colony stimulating factor; secretion signal; meg-CSF; stability; proteolytic cleavage; adhesion; alternative splicing.

Synthetic.

Key Location/Qualifiers

1..26 /label= Exon_I

26..67 /label= Exon_II

67..107 /label= Exon_III

107..157 /label= Exon_IV

157..200 /label= Exon_V

200..1141 /label= Exon_VI

1141..1166 /label= Exon_VII

1166..1212 /label= Exon_VIII

1212..1266 /label= Exon_IX

1266..1331 /label= Exon_X

1331..1373 /label= Exon_XI

1373..1404 /label= Exon_XII

MO9213075-A.

06-AUG-1992.

17-JAN-1992; 92MO-US00433.

18-JAN-1991; 91US-0643502.

10-SEP-1991; 91US-0757022.

(GENY) GENETICS INST INC.

Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K;

antibodies and research use -

Genetic intervention provides the protein and coding sequences of novel proteins from a variety of organisms, including human, dog, cat, horse, cow, pig, hamster, monkey, meacque, yeast, bacteria, fruit fly, sea urchin, and tomato. These were derived from expressed sequence tags (ESTs) from the organism of interest. They can be used in diagnostics, forensic, gene mapping, identification of mutations, to assess biodiversity and for nutritional purposes. The present sequence is a protein of the invention.

Query Match	100.0%;	Score 5698;	DB 22;	Length 1299;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1064; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

Matches 1064; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

[illegible]

ID AAR26049 standard; Protein; 1404 AA.

DT 02-FEB-1993 (first entry)
XX

XX
KM
XX
stability; proteolytic cleavage; adhesion; alternative splicing.
megakaryocyte colony stimulating factor; secretin signal; meg-CSF
megakaryocyte colony stimulating factor; secretin signal; meg-CSF

EH	key	Location/Qualifiers
ET	Region	1 35

FT	Exon	Exon_II
67	107	107

ET	Position	/label = Exon_IV
ET	153 000	

Exon	Label	Exon_VI
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

FT	Region	/rabe1 - Exon_VIII
1213	1266	

Region	1331..1373
--------	------------

AA
PN
W09213075-A

PF 17-JAN-1992; 92WO-US00433.



RA (GENE) GENETICS INST INC.

PI Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K;

XX WPI: 1992-284660/34.

DR N-PSDB: MA027223.

PT New human mega-karyocyte stimulating factors - for treating
PT immune deficiencies, cancer, exposure to radiation or drugs,
XX bacterial and viral infections, etc.

PS Claim 1, 2 and 3: Fig 1: 87pp: English.

CC The sequence given is a full length translation from the megakaryocyte
CC stimulating factor (MSF) precursor. The sequence covered by exons II,
CC III and IV encodes megakaryocyte stimulating factor (MSF). This
CC sequence is modified by the addition of an N-terminal sequence encoding
CC a secretory leader, an initiating methionine preceding exon II and a
CC terminating codon following exon IV. The cDNA sequence given contains
CC (meg-CSF). Exon I contains the initiating methionine, and encodes a
CC classical mammalian protein secretion signal sequence. The sequence
CC encoding the original meg-CSF includes exons II-IV and is thought to
CC terminate in the region between amino acid residues 134 - 147. The
CC primary transcript of this gene may be cleaved in different ways to
CC yield a family of mRNA's each encoding a different MSF protein. Exons
CC V and VI are thought to be related to the activity of the factor and
CC are also implicated in the stability, folding and processing of the
CC molecule. These exons are also thought to play a role in the observed
CC synergy of MSF with other cytokines. Exons V - XII are believed to be
CC implicated in the processing or folding of the appropriate structure of
CC the resulting factor, i.e. one or more of these exons may contain
CC sequences which direct proteolytic cleavage, adhesion, organization of
CC the cellular matrix or extracellular matrix processing. Both naturally
CC occurring and non-naturally occurring MSF's may be characterized by
CC various combinations of alternatively spliced exons from this sequence,
CC with the exons spliced together in differing orders to form different
CC members of the MSF family.

XX Sequence 1404 AA;

Query Match 100.08; Score 5698; DB 13; Length 1404;
Best Local Similarity 100.08; Pred. No. 0;
Matches 1064; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDNKKKRTKRRKPPVYVDSGLDNGDFKVTTPDSTQHNKVSPPKITTAKPLN 60
DB 200 VDNKKKRTKRRKPPVYVDSGLDNGDFKVTTPDSTQHNKVSPPKITTAKPLN 60
QY 61 PRSSLPNSDTSKETSIVNKKETVETKETTNNKOTSDGKETSAREQSIETSAK 120
DB 260 PRSSLPNSDTSKETSIVNKKETVETKETTNNKOTSDGKETSAREQSIETSAK 120
QY 121 DLAPTSKYLAKRPFAETTTGSPALTTPKEPPTTPKPASTTPKEPPTTTSAPTPPK 180
DB 320 DLAPTSKYLAKRPFAETTTGSPALTTPKEPPTTPKPASTTPKEPPTTTSAPTPPK 180
QY 330 DLAPTSKYLAKRPFAETTTGSPALTTPKEPPTTPKPASTTPKEPPTTTSAPTPPK 180
DB 379 DLAPTSKYLAKRPFAETTTGSPALTTPKEPPTTPKPASTTPKEPPTTTSAPTPPK 180
QY 131 EPAPPTTSAPTPKEPAPPTTPKEPAPPTTPKEPAPPTTPKEPAPPTTPKEPAPPTT 240
DB 380 EPAPPTTSAPTPKEPAPPTTPKEPAPPTTPKEPAPPTTPKEPAPPTTPKEPAPPTT 240
QY 241 PKKPAPTPKKPAPTTPKEPPTTPKEPAPPTTPKEPAPPTTPKEPAPPTTPKEPAP 300
DB 440 PKKPAPTPKKPAPTTPKEPPTTPKEPAPPTTPKEPAPPTTPKEPAPPTTPKEPAP 300
QY 301 PTPKEPAPPTTPKEPPTTPKEPAPPTTPKEPAPPTTPKEPAPPTTPKEPAPPTTPKE 360
DB 500 PTPKEPAPPTTPKEPPTTPKEPAPPTTPKEPAPPTTPKEPAPPTTPKEPAPPTTPKE 360
QY 361 PAPTTPKEPAPPTTPKEPAPPTTPKEPAPPTTPKEPAPPTTPKEPAPPTTPKEPAP 420
DB 560 PAPTTPKEPAPPTTPKEPAPPTTPKEPAPPTTPKEPAPPTTPKEPAPPTTPKEPAP 420

QY 421 TPKKLTPTEPKLAPPTTPKEPAPPTTPKEPAPPTTPKEPAPPTTPKEPAPPTTPKE 480
DB 620 TPKKLTPTEPKLAPPTTPKEPAPPTTPKEPAPPTTPKEPAPPTTPKEPAPPTTPKE 480
QY 481 PAPTTPKEPAPPTTPKEPAPPTTPKEPAPPTTPKEPAPPTTPKEPAPPTTPKEPAP 540
DB 680 PAPTTPKEPAPPTTPKEPAPPTTPKEPAPPTTPKEPAPPTTPKEPAPPTTPKEPAP 540
QY 541 EPNSTSDKPAPTTPKGAATTPKEPAPPTTPKEPAPPTTPKEPAPPTTPKEPAPPTT 600
DB 740 EPNSTSDKPAPTTPKGAATTPKEPAPPTTPKEPAPPTTPKEPAPPTTPKEPAPPTT 600
QY 601 KELAPPTTPGPTSTSDKPAPTTPKEPAPPTTPKEPAPPTTPKEPAPPTTPKEPAPPT 660
DB 800 KELAPPTTPGPTSTSDKPAPTTPKEPAPPTTPKEPAPPTTPKEPAPPTTPKEPAPPT 660
QY 661 PPTTKEPPTTHKSPDSPELSAEPPTPALENSPEKPGVPTTKTAAATPEMTTAKDPT 720
DB 860 PPTTKEPPTTHKSPDSPELSAEPPTPALENSPEKPGVPTTKTAAATPEMTTAKDPT 720
QY 721 TPBDLRTPEPTTAPKMTKEATTEKTTESKITAATTTQVSTTODTTPPKITTLKTT 780
DB 920 TPBDLRTPEPTTAPKMTKEATTEKTTESKITAATTTQVSTTODTTPPKITTLKTT 780
QY 781 TLAPKVTYTKKITTETTEIMNRPDEETAKPKDRATNSKATTPKOKTKAPKPKTSKKKT 840
DB 980 TLAPKVTYTKKITTETTEIMNRPDEETAKPKDRATNSKATTPKOKTKAPKPKTSKKKT 840
QY 841 MPRAVKKPTTPPKMTSTMBELNPTSRISIAAMLOTTRNQTNPNSKLVENPKSEADAG 900
DB 1040 MPRAVKKPTTPPKMTSTMBELNPTSRISIAAMLOTTRNQTNPNSKLVENPKSEADAG 900
QY 901 AEGETPHMLLPHFMPVPTPDMDYLPRVNOGIINMLSDENINCKGVYDGLTLRN 960
DB 1100 AEGETPHMLLPHFMPVPTPDMDYLPRVNOGIINMLSDENINCKGVYDGLTLRN 960
QY 961 GTLVAFRGHFWMLSPSPSPARRITEWGIIPSIDVTFRCNCEGTFEFGKSOVMRF 1020
DB 1160 GTLVAFRGHFWMLSPSPSPARRITEWGIIPSIDVTFRCNCEGTFEFGKSOVMRF 1020
QY 1021 TNDIKDGYPRPKFGGGLGQIVAAALSTAKYNNPESYFEK 1064
DB 1220 TNDIKDGYPRPKFGGGLGQIVAAALSTAKYNNPESYFEK 1064
RESULT 3
AAB60568
ID AAB60568 standard; Protein: 1404 AA.
AC AAB60568;
XX 27-APR-2001 (first entry)
DT 27-APR-2001 (first entry)
XX Human megakaryocyte stimulating factor (MSF, CACP).
DE Human megakaryocyte stimulating factor (MSF, CACP).
XX Human, CACP protein; campylocactyly-arthropathy-coxa vara-pericarditis;
KW MSF; megakaryocyte stimulating factor; synovial lubricant;
KW chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;
XX antiarthritic.
OS Homo sapiens.
XX Homo sapiens.
XX WO200107068-A.
PN WO200107068-A.
XX 01-FEB-2001.
PD 01-FEB-2001.
XX 21-JUL-2000; 2000MO-US20002.
PF 21-JUL-2000; 2000MO-US20002.
XX 23-JUL-1999; 99US-0145328.
PR 23-JUL-1999; 99US-0145328.
XX 19-JUL-2000; 2000US-0145328.
XX 19-JUL-2000; 2000US-0145328.
PA (UYCA-) UNIT CASE WESTERN RESERVE.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:15:40 ; Search time 107.17 Seconds

(without alignments)
953.822 Million cell updates/sec

Title: US-09-556-246-1_COPY_25_1404

Perfect score: 7410
Sequence: 1 QDSSCAGRCGEGYSRDATC.....AARITRSCGTLSKWTNCP 1380

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

A_Genseq_1101:*

1: /SIDSR/gcgdata/genseq/genseq/AA1980.DAT:*
2: /SIDSR/gcgdata/genseq/genseq/AA1981.DAT:*
3: /SIDSR/gcgdata/genseq/genseq/AA1982.DAT:*
4: /SIDSR/gcgdata/genseq/genseq/AA1983.DAT:*
5: /SIDSR/gcgdata/genseq/genseq/AA1984.DAT:*
6: /SIDSR/gcgdata/genseq/genseq/AA1985.DAT:*
7: /SIDSR/gcgdata/genseq/genseq/AA1986.DAT:*
8: /SIDSR/gcgdata/genseq/genseq/AA1987.DAT:*
9: /SIDSR/gcgdata/genseq/genseq/AA1988.DAT:*
10: /SIDSR/gcgdata/genseq/genseq/AA1989.DAT:*
11: /SIDSR/gcgdata/genseq/genseq/AA1990.DAT:*
12: /SIDSR/gcgdata/genseq/genseq/AA1991.DAT:*
13: /SIDSR/gcgdata/genseq/genseq/AA1992.DAT:*
14: /SIDSR/gcgdata/genseq/genseq/AA1993.DAT:*
15: /SIDSR/gcgdata/genseq/genseq/AA1994.DAT:*
16: /SIDSR/gcgdata/genseq/genseq/AA1995.DAT:*
17: /SIDSR/gcgdata/genseq/genseq/AA1996.DAT:*
18: /SIDSR/gcgdata/genseq/genseq/AA1997.DAT:*
19: /SIDSR/gcgdata/genseq/genseq/AA1998.DAT:*
20: /SIDSR/gcgdata/genseq/genseq/AA1999.DAT:*
21: /SIDSR/gcgdata/genseq/genseq/AA2000.DAT:*
22: /SIDSR/gcgdata/genseq/genseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7410	100.0	1404	13	MSF precursor. Sy
2	7410	100.0	1404	22	Human megakaryocyt
3	7410	100.0	1404	22	Human megakaryocyt
4	6834	92.2	1299	22	Human EST encoded
5	3484	47.0	902	22	Human MSF-derived
6	2081	28.1	452	16	Human megakaryocyt
7	1707.5	23.0	472	22	Bovine MSF ortholo
8	1188	16.0	5179	22	AA6204516
9	981	13.2	763	21	AA638942
10	950	12.8	1664	19	AA643106
11	805	10.9	1325	22	AA603645

12	752	10.1	182	12	AAR10872
13	715	9.6	763	18	AAW31852
14	645.5	8.7	4412	21	AAV53666
15	613	8.3	111	13	AAR26050
16	612	8.3	572	18	AAW31855
17	555.5	7.5	844	7	AA60570
18	542	7.3	807	21	AAV54467
19	520.5	7.0	788	21	AAV54466
20	513	6.9	1837	21	AAW11726
21	506.5	6.8	744	9	AAW82875
22	498.5	6.7	1721	21	AAW11727
23	498	6.7	1721	19	AAW48299
24	496.5	6.7	2971	21	AAW41231
25	489.5	6.6	2972	22	AAW50363
26	489.5	6.6	3118	22	AAW50362
27	488.5	6.6	826	13	AAR26042
28	488	6.6	617	22	AAW16458
29	488	6.6	617	22	AAW04187
30	488	6.6	957	21	AAV59288
31	488	6.6	957	22	AAW24513
32	485	6.5	1127	22	AAW55541
33	476.5	6.4	652	9	AAW82974
34	467.5	6.3	511	22	AAW14883
35	467.5	6.3	511	22	AAW27312
36	467.5	6.3	511	22	AAW02607
37	454.5	6.1	1012	20	AAW17406
38	454.5	6.1	2819	22	AAW35408
39	450.5	6.1	378	12	AAW14160
40	448	6.0	1237	21	AAW81609
41	447.5	6.0	3266	21	AAW42491
42	446.5	6.0	378	12	AAW14162
43	446.5	6.0	750	20	AAW05477
44	446.5	6.0	2665	22	AAW14533
45	446.5	6.0	2665	22	AAW26950

ALIGNMENTS

RESULT 1	
ID AAR26049	standard; Protein; 1404 AA.
AC AAR26049;	
DT 02-FEB-1993 (first entry)	
DE	
XX MSF precursor.	
XX	
KW Megakaryocyte colony stimulating factor; meg-CSF;	
KW stability; proteolytic cleavage; adhesion; alternative splicing.	
XX	
OS Synthetic.	
XX	
FH Key	Location/Qualifiers
FT Region	1..26
FT Region	/label- Exon_I
FT Region	26..67
FT Region	/label- Exon_II
FT Region	67..107
FT Region	/label- Exon_III
FT Region	107..157
FT Region	/label- Exon_IV
FT Region	157..200
FT Region	/label- Exon_V
FT Region	200..1141
FT Region	/label- Exon_VI
FT Region	1411..1166
FT Region	/label- Exon_VII
FT Region	1166..1212
FT Region	/label- Exon_VIII
FT Region	1213..1266
FT Region	/label- Exon_IX

Protein encoded by
Mycobacterium tube
Sequence g1/01742
MSF-K130.. Synthet
Mycobacterium tube
Sequence of the Fa
Aml1 acid sequenc
Amino acid sequenc
Cryptosporidium pa
Bioadhesive precu
Portion of Cryptos
Cryptosporidium pa
Human OREX ORF995
Human SRCAP. Homo
P. yoelii SSP2 ant
Peptide #2892 enco
Peptide #2869 enco
Human MUC11 polype
9900P predicted am
Human protein sequ
Bioadhesive precu
Peptide #1317 enco
Peptide #1349 enco
Peptide #1289 enco
Human atrophin-1 r
Human 07C627 gene
PRP 378. Trilicium
Streptococcus pneu
Human OREX ORF2355
PRP encoded by clo
C. albicans Rbt1 p
Peptide #867 enco
Peptide #867 enco

FT	Region	1266..1331
FT	/label= Exon_X	
FT	Region	1331..1373
FT	/label= Exon_XI	
FT	Region	1373..1404
FT	/label= Exon_XII	
XX	WC09213075-A.	
PN		
PD	06-AUG-1992.	
XX		
PE	17-JAN-1992; 92MO-US00433.	
XX		
PR	18-JAN-1991; 91US-0643502.	
PR	10-SEP-1991; 91US-0757022.	
PA	(GENM) GENETICS INST INC.	
XX		
PI	Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K;	
XX		
DR	WPI; 1992-284660/34.	
XX	N-PDSB; AAQ27223.	
XX		
PT	New human mega-karyocyte stimulating factors - for treating	
PT	immune deficiencies, cancer, exposure to radiation or drugs,	
PT	bacterial and viral infections, etc.	
PS	Claim 1, 2 and 3; Fig 1; 87pp; English.	
CC		
CC	The sequence given is a full length translation from the megakaryocyte	
CC	stimulating factor (MSF) precursor. The sequence covered by exons II,	
CC	III and IV encodes megakaryocyte stimulating factor (MSF). This	
CC	sequence is modified by the addition of an N-terminal sequence encoding	
CC	a secretory leader, an initiating methionine proceeding exon II and a	
CC	terminating codon following exon IV. The cDNA sequence given contains	
CC	sequences derived from human megakaryocyte colony stimulating factor	
CC	(meg-CSF). Exon I contains the initiating methionine, and encodes a	
CC	classical mammalian protein secretion signal sequence. The sequence	
CC	encoding the original meg-CSF includes exons II-IV and is thought to	
CC	terminate in the region between amino acid residues 134 - 147. The	
CC	primary transcript of this gene may be cleaved in different ways to	
CC	yield a family of mRNA's each encoding a different MSF protein. Exons	
CC	V and VI are thought to be related to the activity of the factor and	
CC	are also implicated in the stability, folding and processing of the	
CC	molecule. These exons are also thought to play a role in the observed	
CC	synergy of MSF with other cytokines. Exons V - XII are believed to be	
CC	implicated in the processing or folding of the appropriate structure of	
CC	the resulting factor, ie. one or more of these exons may contain	
CC	sequences which direct proteolytic cleavage, adhesion, organisation of	
CC	the cellular matrix or extracellular matrix processing. Both naturally	
CC	occurring and non-naturally occurring MSF's may be characterised by	
CC	various combinations of alternatively spliced exons from this sequence,	
CC	with the exons spliced together in differing orders to form different	
CC	members of the MSF family.	
XX		
SQ	Sequence 1404 AA:	
	Query Match 100.0%; Score 7410; DB 13; Length 1404;	
	Best Local Similarity 100.0%; Pred. No. 0;	
	Matches 1380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	1 ODLSSCAGRCGEYGRDATCNCDYNOCHYMECCPDFRVCRTALSLCKGRGFESFEGRGEC 60	
DB		
DB	25 qdlsscagrcgeygrdatcncdyncqnhymecpddfrvctaalsckgyrcfcsfergrrc 84	
OY	61 DCDACCKRYDKCPCDYESFCAEVHNPTSPSSKRAPPSGASOTIKSTRKSPKPNKKX 120	
DB		
DB	85 dcdagckkydkccpdyesfcaevehnpsspskrapppsgaagtlkstrpkipnkkk 144	
OY	121 TKRYTESSELTREHSVSNQSSSSSSSSSSSTIWKIKSKSKSANRELQKTLKYDKM 180	
DB		
DB	145 tkrytesseeltrehsvenqsssssssssstlwkikssksaanrelqkklkydkm 204	

Qy	181	KNKTKKPKPKPPVVDKAGSGJDNDDFKYTPPDSTJOHKNKYSTSKITTAKPINRPSL	240
Db	205	KnrlkkkcltpkppvvdaagsldngdIfkvtcpdstcltnhkvstspkltcaKpInprsl	264
Qy	241	PNNSOTSKETSLTVNKKETTVERTKETTNTNKKONSTOGKETTTSKKEQSLSEKTSAKDLAPT	300
Db	265	pnnsdtskcscltnvhkettvelkctlnqksgtqkctsaekqslsekUsaaklapt	324
Qy	301	SKVLAKPPTKASTTTPKGPALPTTKKEPTPTPKKEPASTTKEEPTPTTKASAPTTPKKEBAPT	360
Db	325	skvLkKpKpkaecttkkgpalltckpkeptctpkpkepastctpkkeptcltkasptckpkeapc	384
Qy	361	TKKSAPTTPKKEBAPTTPTKKEBAPTTPKKEBAPTTPTKKEBAPTTPTKSAPTTPKKEBAPTTPKKA	420
Db	385	tkksapctpkpkeapcttkkcapcttpkpkpkeapcttkkcapcttkksapcttpkpkpkeapcttkkpa	444
Qy	421	PTTPKKEBAPTTPKKEPTPTTPKKEBAPTTPKKEBAPTTPKKEBAPTTPKKAAPTTPKKEBAPTTPK	480
Db	445	pttpkKpaptctpkkeptctpkpkpkeapcttpkpkpkeapcttkkKpaptctpkpkeapcttpk	504
Qy	481	BEAPATTPKPSPTTPKKEBAPTTPTKSAPTTPTKKEBAPTTPTKSAPTTPKKEBAPTTPKKEBAPT	540
Db	505	epapcttkkspsctpkpkeapcttkksapcttkkcapcttkksapcttpkpkpkeapcttkkcapct	564
Qy	541	PKEPAPTTPKKEBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPTTPPKKL	600
Db	565	pKpEapcttpkKpaptctpkpkeapcttpkpkpkeapcttkkKpaptckapkeapcttpkpetaptcpkkl	624
Qy	601	PTTPPKKLAPTTPKKEBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPTTPKKAAPTTPKKEBAPTTP	660
Db	625	tpctpkEapcttpkKpaptctpkpkeapcttpkEapcttpkEapcttpkkaapnctpkpkeapct	684
Qy	661	PKKEBAPTTPKKEBAPTTPKKEBAPTTPKKGATPTTKKEBAPTTPKKAKEBAPTTPKKEBAPT	720
Db	685	pKpEapcttpkKpaptctpkpkeapcttpkEapcttkkKpaptckapkeapcttpkEapcttkkepst	744
Qy	721	TSDKBAPTTPKKGATPTTPKKEBAPTTPKKEBAPTTPKKGATPTTKKEBAPTTPKKAKEBAPTTP	780
Db	745	tsdkpEapcttpkKpaptctpkpkeapcttpkEapcttkkKpaptckapkeapcttpkKpapelap	804
Qy	781	TTTTKGPSTTSDKRAPTTTKETAAPTTPKKEBAPTTPKKAAPTTPETPTPTTSBVSSTPTTK	840
Db	805	tttkgptstsdKpaptctpkketaptctpkpkeapcttpkKpaptctpEppctsevspttk	864
Qy	841	EPPTTKSPDESPELSABPTPKALNSKPEEVPPTTKPKAARPEMTTAAKOTTERDL	900
Db	865	eppttkspdesPElsabptPkalsnsKpEevpttkPkAARPEMTTAAKOTTERdl	924
Qy	901	RTTPENTTAPKMTKETATTEKTTESKITTATTOYSTTODTPEKITTLLKTTTLTAPK	960
Db	925	rttpeTtaapKmtketatttEktteskIttatctqvtstctgdtptckltcltkttclapK	984
Qy	961	VTTTKKTTTTTETIMNKPBEETAKPKDRAVNSKATTPKQKFTAKPKKFTSTKKPKTMRVR	1020
Db	985	vtttkKtlttTetImnKpEeetAkPkdravnsKattPkqKftKpKclKpKctskKpkltmPrv	1044
Qy	1021	KKPTTTPPKMNSTMEALNPTSSIIAAMLOTTTRPNOTPNRSKILVENVPKSEBDAGGAGET	1080
Db	1045	kKptcttpKmtsmelnpTssIIaamLotttrPnqtpnKsilvenvpKseedaggaaget	1104
Qy	1081	PHMLRBPHEMPEVTPDMOVLPEVPNOGIIINMLSPETNICNGKPDVAGJTLLENGTLVA	1140
Db	1105	phmlrlbpHemPeVtpdmovLpevpNoGiiInmlspEtNicngKpdvAgjtlleNgtlva	1164
Qy	1141	FRGHYFMWLSPESPSPARRITEWGLIPSDIDVTTRCNCNGKTFEFKDSQYWFMTNDIK	1200
Db	1165	frghyfmwlsPespSparriteWglIpsdIdvttrCncngKtfEfkdsqyWfMtndik	1224
Qy	1201	DAGYKPRIRKGFGLGQIYVALSTAKYKMWPSVYFRFKGGSIQOYIYKOEPPQKCPGR	1260
Db	1225	dagYkPrIrKgfGlGqIyvalStakYkMwpsvYfrfkGgsIqoYIyKoePpQkCpgr	1284

QY 1261 RPALNTPVYGEHTOYRRRERBAIGSPQTHTRIOYSPARLAKODKGVLAHNEKVISIMR 1320
DB 1285 RPALHPYGEHTYRRRERBAIGSPQTHTRIOYSPARLAKODKGVLAHNEKVISIMR 1344
QY 1321 GUPNVTSAISLIPNIRKPDGYDYAFSKNDQYINIVPSTARITRSGQTLSTKVMWCP 1380
DB 1345 GUPNVTSAISLIPNIRKPDGYDYAFSKNDQYINIVPSTARITRSGQTLSTKVMWCP 1404

RESULT 2

AA060568 standard; Protein: 1404 AA.

AA060568:

27-APR-2001 (first entry)

Human megakaryocyte stimulating factor (MSF, CACP).

Human CACP protein, campodactyl-arthropathy-coxa vara-pericarditis;

MSF, megakaryocyte stimulating factor; synovial lubricant;

chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;

antiarthritic.

Homo sapiens.

MO200107068-A1.

01-FEB-2001.

21-JUL-2000; 2000MO-US20002.

23-JUL-1999; 199US-0145328.

19-JUL-2000; 2000US-0145328.

(UTCA-) UNIV CASE WESTERN RESERVE.

Warman MI.

WPI: 2001-182721/18.

New composition comprising the campodactyl-arthropathy-coxa

vara-pericarditis protein in combination with an anesthetic, useful for

treating osteoarthritis or as lubricants of tissue and joints

Example 1: Page -: 34pp: English.

The invention relates to a method of treating osteoarthritis via the
administration of a composition comprising the campodactyl-arthropathy-
coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.
The composition may further comprise a local anesthetic. The composition
of the invention may be administered via intra-articular or intravenous
injection. The human CACP protein is identified in the invention as
being megakaryocyte stimulating factor (MSF). The gene encoding
CACP protein (MSF) is located on chromosome 1q25-31, and mutations in
this gene are responsible for the heritable disorder campodactyl-
arthropathy-coxa vara-pericarditis, in which patients have synovial
hyperplasia without evidence of inflammation. CACP protein (MSF)
acts as a synovium lubricant, and can be used to lubricate tissue and
joints in the treatment of osteoarthritis. The composition may be
applied to reduce the symptoms of osteoarthritis (e.g., joint pain,
loss of range of movement or joint damage). The present sequence
represents human megakaryocyte stimulating factor (MSF, CACP protein).
Note: This sequence is not given in its entirety in Figure 4 of the
specification, although a GenBank accession number was given. This
sequence was therefore obtained from GenBank (U070316).

Sequence 1404 AA:

Query Match 100.0%; Score 7410; DB 2; Length 1404;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1360; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY ODSSCAGRGEGYSRDATCNCDVNOHMECCPDFKRVCTAELSCGRPFESPERREC 60
DB 25 Gdlssccagrcgsgysrdatcncdvnohmeccpdfkrrvctaelscgrpfesfegrec 84
QY DCAOACKKDYKCCPDYEFSCAEVHNPTSPSSKAPPGASQTIKSTKRSPPPKR 120
DB 85 ddaqckkykccpdysfcaevhnptspsskappgsaqtllsktrspbkprk 144
QY 121 TKVIESEETTEHSVSENDESSSSSSSSSTIMKSSKNSAANRELQKLVKDNK 180
DB 145 tkvleeeelteesvsenadesssssssstlwklsksnaanrelqkllkvdkn 204
QY 181 KNRRTKKRPKPPVYDAGSGLDNGDFKVTTPPTSTQHNKYSTSKRTTAKINRPSL 240
DB 205 knrtkkpckppvdeagsgldngdfkvtctpdtstqhnkvsstltpklnprpsl 264
QY 241 PPSNDTSKETSLSLVNKEETVETKETTINKQSTDGKEKTSKETSQSIKTSKDLAPT 300
DB 265 ppsndtsketslslvnketvettinkqstdgketsketsketsqslktsakdlapt 324
QY 301 SKVLAKTPKAEHTTGGPALTTPKEPTTPKPASTTPKEPTTIKSAPTTPKPA 360
DB 325 skvlaktpkaehttggpalttpkepttpkpaasttpkepttiiksapttpkpa 384
QY 361 TTKSAPTPKEAPTPTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPA 420
DB 385 tksapttpkeaptptpkpaattpkpaattpkpaattpkpaattpkpaattpkpa 444
QY 421 PTPKPAATTPKPAATTPTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPK 480
DB 445 ptpkpaattpkpaattptpkpaattpkpaattpkpaattpkpaattpkpaattpk 504
QY 481 EPAPTTKPSPTTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTP 540
DB 505 epapttkpspttpkpaattpkpaattpkpaattpkpaattpkpaattpkpaattp 564
QY 541 PKPAPTPPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPA 600
DB 565 pkpaptppkpaattpkpaattpkpaattpkpaattpkpaattpkpaattpkpa 624
QY 601 TPTTPKLAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTP 660
DB 625 tpttpklaattpkpaattpkpaattpkpaattpkpaattpkpaattpkpaattp 684
QY 661 PKPAPTPPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPA 720
DB 685 pkpaptppkpaattpkpaattpkpaattpkpaattpkpaattpkpaattpkpa 744
QY 721 TSDKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPK 780
DB 745 tsdkpaattpkpaattpkpaattpkpaattpkpaattpkpaattpkpaattpk 804
QY 781 TTTKGPSTTSKAPPTTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPK 840
DB 805 tttkgpsttskappttpkpaattpkpaattpkpaattpkpaattpkpaattpk 864
QY 841 EPTTIHKSDESTPELSAPPKALENSPKSGVTTTPPAATTKBMTTAKDKTTERDL 900
DB 865 epttihkstpeelsapppkalenspsksgvtttppaattkbmtttakdktterdl 924
QY 901 RTTPPTTAAAPKMTKETAATTEKTESKITATTTQVSTTODTTPFKITTKTTTAPK 960
DB 925 rtpppttaapkmteattekteskitattttqvsttoddttppfkittktttap 984
QY 961 VTTTKKTTTTEIMKPEETAKPKDRATNSKATTPKPKPTPAKPKPTSTKPKTTPRV 1020
DB 985 vttkktttteimkpeetakpkdratnskattpkpkptpaakpkptsttkpktp 1044
QY 1021 KPTTTPPKMTSTWPELNPSTIAEAMLOTTRPNQHNKSTLVNPKSDDAGABSEET 1080
DB 1045 kpttppkmtstwpelnpstiaeamlottrpnqhnkstlvnpskddagabseet 1104

